

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
19 July 2001 (19.07.2001)

PCT

(10) International Publication Number
WO 01/51636 A2

(51) International Patent Classification: **C12N 15/12.**
C07K 14/47, 16/18, A61K 38/17, A01K 67/027, G01N
33/50, 33/68, C13Q 1/68

TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose,
CA 95118 (US). BAUGHN, Mariah, R. [US/US]; 14244
Santiago Road, San Leandro, CA 94577 (US). LU, Dyoung,
Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123
(US). YANG, Junming [CN/US]; 7125 Bark Lane, San
Jose, CA 95129 (US). BURFORD, Neil [GB/US]; 105
Wildwood Circle, Durham, CT 06422 (US). AU-YOUNG,
Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA
94005 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley
Avenue #3, Sunnyvale, CA 94086 (US).

(21) International Application Number: PCT/US01/00895

(22) International Filing Date: 10 January 2001 (10.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(74) Agents: HAMLET-COX, Diana et al.: Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(30) Priority Data:

60/176,113	14 January 2000 (14.01.2000)	US
60/177,733	21 January 2000 (21.01.2000)	US
60/178,832	28 January 2000 (28.01.2000)	US
60/179,774	2 February 2000 (02.02.2000)	US
60/186,792	3 March 2000 (03.03.2000)	US

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, L, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(71) Applicant (for all designated States except US): INCYTE
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo
Alto, CA 94304 (US).

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (for US only): LAL, Preeti
[IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US).
BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain
View, CA 94043 (US). PATTERSON, Chandra
[US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025
(US). WALIA, Narinder, K. [US/US]; 890 Davis Street
#205, San Leandro, CA 94577 (US). NGUYEN, Danniel,
B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118
(US). YUE, Henry [US/US]; 826 Lois Avenue, Sunny-
vale, CA 94087 (US). KHAN, Farrah, A. [IN/US]; 333
Liscuela Avenue #221, Mountain View, CA 94040 (US).

Published:

without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 01/51636 A2

(54) Title: SECRETED PROTEINS

(57) Abstract: The invention provides human secreted proteins (SECP) and polynucleotides which identify and encode SECP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of SECP.

BEST AVAILABLE COPY

SECRETED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of secreted proteins and to the
5 use of these sequences in the diagnosis, treatment, and prevention of cell proliferative,
autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders, and in the
assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid
sequences of secreted proteins.

10

BACKGROUND OF THE INVENTION

Protein transport and secretion are essential for cellular function. Protein transport is mediated
by a signal peptide located at the amino terminus of the protein to be transported or secreted. The
signal peptide is comprised of about ten to twenty hydrophobic amino acids which target the nascent
protein from the ribosome to a particular membrane bound compartment such as the endoplasmic
15 reticulum (ER). Proteins targeted to the ER may either proceed through the secretory pathway or
remain in any of the secretory organelles such as the ER, Golgi apparatus, or lysosomes. Proteins that
transit through the secretory pathway are either secreted into the extracellular space or retained in the
plasma membrane. Proteins that are retained in the plasma membrane contain one or more
transmembrane domains, each comprised of about 20 hydrophobic amino acid residues. Secreted
20 proteins are generally synthesized as inactive precursors that are activated by post-translational
processing events during transit through the secretory pathway. Such events include glycosylation,
proteolysis, and removal of the signal peptide by a signal peptidase. Other events that may occur
during protein transport include chaperone-dependent unfolding and folding of the nascent protein and
interaction of the protein with a receptor or pore complex. Examples of secreted proteins with amino
25 terminal signal peptides are discussed below and include proteins with important roles in cell-to-cell
signaling. Such proteins include transmembrane receptors and cell surface markers, extracellular
matrix molecules, cytokines, hormones, growth and differentiation factors, enzymes, neuropeptides, and
vasomediators. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of The Cell, Garland
Publishing, New York, NY, pp. 557-560, 582-592.)

30 Cell surface markers include cell surface antigens identified on leukocytic cells of the immune
system. These antigens have been identified using systematic, monoclonal antibody (mAb)-based "shot
gun" techniques. These techniques have resulted in the production of hundreds of mAbs directed
against unknown cell surface leukocytic antigens. These antigens have been grouped into "clusters of
differentiation" based on common immunocytochemical localization patterns in various differentiated

and undifferentiated leukocytic cell types. Antigens in a given cluster are presumed to identify a single cell surface protein and are assigned a "cluster of differentiation" or "CD" designation. Some of the genes encoding proteins identified by CD antigens have been cloned and verified by standard molecular biology techniques. CD antigens have been characterized as both transmembrane proteins and cell surface proteins anchored to the plasma membrane via covalent attachment to fatty acid-containing glycolipids such as glycosylphosphatidylinositol (GPI). (Reviewed in Barclay, A. N. et al. (1995) The Leucocyte Antigen Facts Book, Academic Press, San Diego, CA, pp. 17-20.)

Matrix proteins (MPs) are transmembrane and extracellular proteins which function in formation, growth, remodeling, and maintenance of tissues and as important mediators and regulators of the inflammatory response. The expression and balance of MPs may be perturbed by biochemical changes that result from congenital, epigenetic, or infectious diseases. In addition, MPs affect leukocyte migration, proliferation, differentiation, and activation in the immune response. MPs are frequently characterized by the presence of one or more domains which may include collagen-like domains, EGF-like domains, immunoglobulin-like domains, and fibronectin-like domains. In addition, MPs may be heavily glycosylated and may contain an Arginine-Glycine-Aspartate (RGD) tripeptide motif which may play a role in adhesive interactions. MPs include extracellular proteins such as fibronectin, collagen, galectin, vitronectin and its proteolytic derivative somatomedin B; and cell adhesion receptors such as cell adhesion molecules (CAMs), cadherins, and integrins. (Reviewed in Ayad, S. et al. (1994) The Extracellular Matrix Facts Book, Academic Press, San Diego, CA, pp. 2-16; Ruoslahti, E. (1997) *Kidney Int.* 51:1413-1417; Sjaastad, M.D. and Nelson, W.J. (1997) *BioEssays* 19:47-55.)

Hormones are secreted molecules that travel through the circulation and bind to specific receptors on the surface of, or within, target cells. Although they have diverse biochemical compositions and mechanisms of action, hormones can be grouped into two categories. One category includes small lipophilic hormones that diffuse through the plasma membrane of target cells, bind to cytosolic or nuclear receptors, and form a complex that alters gene expression. Examples of these molecules include retinoic acid, thyroxine, and the cholesterol-derived steroid hormones such as progesterone, estrogen, testosterone, cortisol, and aldosterone. The second category includes hydrophilic hormones that function by binding to cell surface receptors that transduce signals across the plasma membrane. Examples of such hormones include amino acid derivatives such as catecholamines (epinephrine, norepinephrine) and histamine, and peptide hormones such as glucagon, insulin, gastrin, secretin, cholecystokinin, adrenocorticotrophic hormone, follicle stimulating hormone, luteinizing hormone, thyroid stimulating hormone, and vasopressin. (See, for example, Lodish et al. (1995) Molecular Cell Biology, Scientific American Books Inc., New York, NY, pp. 856-864.)

Growth and differentiation factors are secreted proteins which function in intercellular communication. Some factors require oligomerization or association with membrane proteins for activity. Complex interactions among these factors and their receptors trigger intracellular signal transduction pathways that stimulate or inhibit cell division, cell differentiation, cell signaling, and cell motility. Most growth and differentiation factors act on cells in their local environment (paracrine signaling). There are three broad classes of growth and differentiation factors. The first class includes the large polypeptide growth factors such as epidermal growth factor, fibroblast growth factor, transforming growth factor, insulin-like growth factor, and platelet-derived growth factor. The second class includes the hematopoietic growth factors such as the colony stimulating factors (CSFs).

5 Hematopoietic growth factors stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and their stem cell precursors. The third class includes small peptide factors such as bombesin, vasopressin, oxytocin, endothelin, transferrin, angiotensin II, vasoactive intestinal peptide, and bradykinin which function as hormones to regulate cellular functions other than proliferation.

10 Growth and differentiation factors play critical roles in neoplastic transformation of cells in vitro and in tumor progression in vivo. Inappropriate expression of growth factors by tumor cells may contribute to vascularization and metastasis of tumors. During hematopoiesis, growth factor misregulation can result in anemias, leukemias, and lymphomas. Certain growth factors such as interferon are cytotoxic to tumor cells both in vivo and in vitro. Moreover, some growth factors and

15 growth factor receptors are related both structurally and functionally to oncoproteins. In addition, growth factors affect transcriptional regulation of both proto-oncogenes and oncosuppressor genes.

20 (Reviewed in Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor, MI, pp. 1-9.)

Neuropeptides and vasomediators (NP/VM) comprise a large family of endogenous signaling molecules. Included in this family are neuropeptides and neuropeptide hormones such as bombesin, neuropeptide Y, neotensin, neuromedin N, melanocortins, opioids, galanin, somatostatin, tachykinins, urotensin II and related peptides involved in smooth muscle stimulation, vasopressin, vasoactive intestinal peptide, and circulatory system-borne signaling molecules such as angiotensin, complement, calcitonin, endothelins, formyl-methionyl peptides, glucagon, cholecystokinin and gastrin. NP/VMs can transduce signals directly, modulate the activity or release of other neurotransmitters and hormones, and act as catalytic enzymes in cascades. The effects of NP/VMs range from extremely brief to long-lasting. (Reviewed in Martin, C.R. et al. (1985) Endocrine Physiology, Oxford University Press, New York, NY, pp. 57-62.)

NP/VMs are involved in numerous neurological and cardiovascular disorders. For example, neuropeptide Y is involved in hypertension, congestive heart failure, affective disorders, and appetite

regulation. Somatostatin inhibits secretion of growth hormone and prolactin in the anterior pituitary, as well as inhibiting secretion in intestine, pancreatic acinar cells, and pancreatic beta-cells. A reduction in somatostatin levels has been reported in Alzheimer's disease and Parkinson's disease. Vasopressin acts in the kidney to increase water and sodium absorption, and in higher concentrations stimulates contraction of vascular smooth muscle, platelet activation, and glycogen breakdown in the liver.

5 Vasopressin and its analogues are used clinically to treat diabetes insipidus. Endothelin and angiotensin are involved in hypertension, and drugs, such as captopril, which reduce plasma levels of angiotensin, are used to reduce blood pressure (Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 194; 252; 284; 55; 111).

10 Neuropeptides have also been shown to have roles in nociception (pain). Vasoactive intestinal peptide appears to play an important role in chronic neuropathic pain. Nociceptin, an endogenous ligand for the opioid receptor-like 1 receptor, is thought to have a predominantly anti-nociceptive effect, and has been shown to have analgesic properties in different animal models of tonic or chronic pain (Dickinson, T. and Fleetwood-Walker, S.M. (1998) Trends Pharmacol. Sci. 19:346-348).

15 Other proteins that contain signal peptides include secreted proteins with enzymatic activity. Such activity includes, for example, oxidoreductase/dehydrogenase activity, transferase activity, hydrolase activity, lyase activity, isomerase activity, or ligase activity. For example, matrix metalloproteinases are secreted hydrolytic enzymes that degrade the extracellular matrix and thus play an important role in tumor metastasis, tissue morphogenesis, and arthritis (Reponen, P. et al. (1995) Dev. Dyn. 202:388-396; Firestein, G.S. (1992) Curr. Opin. Rheumatol. 4:348-354; Ray, J.M. and Stetler-Stevenson, W.G. (1994) Eur. Respir. J. 7:2062-2072; and Mignatti, P. and Rifkin, D.B. (1993) Physiol. Rev. 73:161-195).

20 The discovery of new secreted proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders.

25 The discovery of new secreted proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of secreted proteins.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, secreted proteins, referred to collectively as

"SECP" and individually as "SECP-1," "SECP-2," "SECP-3," "SECP-4," "SECP-5," "SECP-6," "SECP-7," "SECP-8," "SECP-9," and "SECP-10." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-10.

10 The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-10. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:11-20.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group

consisting of SEQ ID NO:1-10. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

5 Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

10 10

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

15 15

20 Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and

25 25

30 30

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of

SEQ ID NO:11-20, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or 5 fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence 10 selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and a pharmaceutically acceptable excipient. In one 15 embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional SECP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an 20 agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino 25 acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of 30 functional SECP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence

selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting 5 antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional SECP, comprising administering to a patient in need of such treatment the composition.

10 The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence selected 15 from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

20 The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of an amino acid sequence 25 selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with 30 the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a

sequence selected from the group consisting of SEQ ID NO:11-20, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said

- 5 method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to 10 a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group 15 consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of 20 a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

25

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank 30 homolog for each polypeptide of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of each polypeptide sequence, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of each polypeptide.

35 Table 4 lists the cDNA and genomic DNA fragments which were used to assemble each

polynucleotide sequence, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for each polynucleotide of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

5 Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood 10 that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," 15 and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings 20 as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, 25 protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"SECP" refers to the amino acid sequences of substantially purified SECP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and 30 human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of SECP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of SECP either by directly interacting with SECP or by acting on components of the biological pathway in which SECP participates.

An "allelic variant" is an alternative form of the gene encoding SECP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to 5 allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding SECP include those sequences with deletions, 10 insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as SECP or a polypeptide with at least one functional characteristic of SECP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding SECP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding SECP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino 15 acid residues which produce a silent change and result in a functionally equivalent SECP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of SECP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and 20 arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, 25 polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. 30 Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of SECP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of SECP either by

directly interacting with SECP or by acting on components of the biological pathway in which SECP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant.

- 5 Antibodies that bind SECP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as

- 20 phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring
- 25 nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic SECP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement,

3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution.

5 Compositions comprising polynucleotide sequences encoding SECP or fragments of SECP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

10 "Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap

15 15 (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
25	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
30	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
35	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp

Val

Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation.

5 (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical

10 modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

15 A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of SECP or the polynucleotide encoding SECP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a

20 fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected

25 from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:11-20 comprises a region of unique polynucleotide sequence that

30 specifically identifies SEQ ID NO:11-20, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:11-20 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:11-20 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:11-20 and the region of SEQ ID NO:11-20 to which the fragment corresponds are routinely

35 determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-10 is encoded by a fragment of SEQ ID NO:11-20. A fragment of SEQ ID NO:1-10 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-10. For example, a fragment of SEQ ID NO:1-10 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-10. The precise length of

5 a fragment of SEQ ID NO:1-10 and the region of SEQ ID NO:1-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full

10 length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a

15 standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence

20 alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is

25 selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several

30 sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2

Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/blast/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 5 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

10 *Gap x drop-off: 50*

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as 15 defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which 20 percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

25 The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of 30 substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and

"diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise 5 comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

10 *Expect: 10*

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for 15 example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

20 "Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely 25 resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized 30 after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among

experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C₆₀t or R₆₀t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of SECP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of SECP which is useful in any of the antibody production methods disclosed herein or known in the art.

5 The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

10 The term "modulate" refers to a change in the activity of SECP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, 15 functional, or immunological properties of SECP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the 15 antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

20 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where 25 necessary to join two protein coding regions, in the same reading frame.

25 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and 30 may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an SECP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of SECP.

30 "Probe" refers to nucleic acid sequences encoding SECP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by

complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

5 Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

10 Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs 15 can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 20 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences 25 and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from 30 their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and

polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to 5 those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the 10 artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a 15 vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA 20 stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear 25 sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing SECP, 30 nucleic acids encoding SECP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or

synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will 5 reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

10 A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, 15 trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well 20 known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an 25 autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the 30 art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria,

cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989),

5 supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 10 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may 15 possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in 20 which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) 25 set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

30 The invention is based on the discovery of new human secreted proteins (SECP), the polynucleotides encoding SECP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide

sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

5 Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

10 Table 3 shows various structural features of each of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS 15 program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

20 Together, Tables 2 and 3 summarize the properties of each polypeptide of the invention, and these properties establish that the claimed polypeptides are secreted proteins. For example, SEQ ID NO:1 is 76% identical to *Mus musculus* lysophosphatidic acid phosphatase (GenBank ID g6635858) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1e-176, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a histidine acid phosphatase domain as 25 determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is an acid phosphatase. SEQ ID NO:2-10 were analyzed and annotated in a similar manner. The algorithms and 30 parameters for the analysis of SEQ ID NO:1-10 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide 5 consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:11-20 or that distinguish between SEQ ID NO:11-20 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA 10 sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

15 The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 1803159H1 is the identification number of an Incyte cDNA sequence, and SINTNOT13 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 60137216V1). Alternatively, the identification numbers in column 5 may refer to 20 GenBank cDNAs or ESTs (e.g., g5523837) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA (e.g., g5523837.v113.gs_3.nt.edit, in which g5523837 is the GenBank identification number of the sequence to which Genscan was applied). The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.)

25 Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was 30 obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to

assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses SECP variants. A preferred SECP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence 5 identity to the SECP amino acid sequence, and which contains at least one functional or structural characteristic of SECP.

The invention also encompasses polynucleotides which encode SECP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20, which encodes SECP. The polynucleotide sequences of 10 SEQ ID NO:11-20, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding SECP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least 15 about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding SECP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-20. 20 Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of SECP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding SECP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the 25 invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring SECP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode SECP and its variants are generally capable of 30 hybridizing to the nucleotide sequence of the naturally occurring SECP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding SECP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons

are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding SECP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

5 The invention also encompasses production of DNA sequences which encode SECP and SECP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding SECP or any fragment thereof.

10 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:11-20 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in

15 "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or 20 combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing 25 system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

30 The nucleic acid sequences encoding SECP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)

Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060).

5 5 Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a

10 10 GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

15 15

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library 20 20 does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments 25 25 30 30 which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode SECP may be cloned in recombinant DNA molecules that direct expression of SECP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent

amino acid sequence may be produced and used to express SECP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter SECP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. Biotechnol.* 17:259-264; and Crameri, A. et al. (1996) *Nat. Biotechnol.* 14:315-319) to alter or improve the biological properties of SECP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding SECP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, SECP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp.55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of SECP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide

having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. 5 (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active SECP, the nucleotide sequences encoding SECP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and 10 inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding SECP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding SECP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding SECP and its initiation codon and upstream regulatory sequences are inserted into 15 the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate 20 for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding SECP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in 25 vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding SECP. These include, but are not limited to, microorganisms such as bacteria transformed 30 with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster

(1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and

5 Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al.

10 (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding SECP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding SECP can be achieved using a

15 multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding SECP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested

20 deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of SECP are needed, e.g. for the production of antibodies, vectors which direct high level expression of SECP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of SECP. A number of vectors

25 containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra;

Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994)

30 *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of SECP. Transcription of sequences encoding SECP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be

used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 5 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding SECP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain 10 infective virus which expresses SECP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of 15 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of SECP in cell lines is preferred. For example, sequences encoding SECP can be transformed into cell 20 lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the 25 introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; 30 Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.)

Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β -glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used.

5 These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the 10 sequence encoding SECP is inserted within a marker gene sequence, transformed cells containing sequences encoding SECP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding SECP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

15 In general, host cells that contain the nucleic acid sequence encoding SECP and that express SECP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

20 Immunological methods for detecting and measuring the expression of SECP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on SECP is preferred, but a competitive binding 25 assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

30 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding SECP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding SECP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to

synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include 5 radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding SECP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence 10 and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode SECP may be designed to contain signal sequences which direct secretion of SECP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the 15 inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture 20 Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid 25 sequences encoding SECP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric SECP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of SECP activity. Heterologous protein and peptide 30 moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion 35 proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site

located between the SECP encoding sequence and the heterologous protein sequence, so that SECP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

5 In a further embodiment of the invention, synthesis of radiolabeled SECP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

10 SECP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to SECP. At least one and up to a plurality of test compounds may be screened for specific binding to SECP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

15 In one embodiment, the compound thus identified is closely related to the natural ligand of SECP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) *Current Protocols in Immunology* 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which SECP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for 20 these compounds involves producing appropriate cells which express SECP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing SECP or cell membrane fractions which contain SECP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either SECP or the compound is analyzed.

25 An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with SECP, either in solution or affixed to a solid support, and detecting the binding of SECP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a 30 labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

35 SECP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of SECP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for SECP

activity, wherein SECP is combined with at least one test compound, and the activity of SECP in the presence of a test compound is compared with the activity of SECP in the absence of the test compound. A change in the activity of SECP in the presence of the test compound is indicative of a compound that modulates the activity of SECP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising SECP under conditions suitable for SECP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of SECP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding SECP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330).

Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding SECP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding SECP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding SECP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal

inbred to overexpress SECP, e.g., by secreting SECP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between 5 regions of SECP and secreted proteins. In addition, the expression of SECP is closely associated with liver tissue. Therefore, SECP appears to play a role in cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders. In the treatment of disorders associated with increased SECP expression or activity, it is desirable to decrease the expression or activity of SECP. In the treatment of disorders associated with decreased SECP expression or activity, it is 10 desirable to increase the expression or activity of SECP.

Therefore, in one embodiment, SECP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue 15 disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, 20 thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, 25 dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid 30 arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive

heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, 25 postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, 30 hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss.

In another embodiment, a vector capable of expressing SECP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased

expression or activity of SECP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified SECP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of SECP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those listed above.

In a further embodiment, an antagonist of SECP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of SECP. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders described above. In one aspect, an antibody which specifically binds SECP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express SECP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding SECP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of SECP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of SECP may be produced using methods which are generally known in the art. In particular, purified SECP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind SECP. Antibodies to SECP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with SECP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to

increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

5 It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to SECP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of SECP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule 10 may be produced.

Monoclonal antibodies to SECP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce SECP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, 25 D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

30 Antibody fragments which contain specific binding sites for SECP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al.

(1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such 5 immunoassays typically involve the measurement of complex formation between SECP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering SECP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques 10 may be used to assess the affinity of antibodies for SECP. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of SECP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple SECP epitopes, represents the average affinity, or avidity, of the antibodies for SECP. The K_a determined for 15 a preparation of monoclonal antibodies, which are monospecific for a particular SECP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10⁹ to 10¹² L/mole are preferred for use in immunoassays in which the SECP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10⁶ to 10⁷ L/mole are preferred for use in immunopurification and similar procedures which ultimately 20 require dissociation of SECP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine 25 the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of SECP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

30 In another embodiment of the invention, the polynucleotides encoding SECP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding SECP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be

designed from various locations along the coding or control regions of sequences encoding SECP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered

5 intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

15 In another embodiment of the invention, polynucleotides encoding SECP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) 20 express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in SECP expression or regulation causes disease, the expression of SECP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in SECP

are treated by constructing mammalian expression vectors encoding SECP and introducing these vectors by mechanical means into SECP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the 5 use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of SECP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), 10 PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). SECP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 15 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous 20 gene encoding SECP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver 25 polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these 30 standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to SECP expression are treated by constructing a retrovirus vector consisting of (i) the 35 polynucleotide encoding SECP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc.*

Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and 5 A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the 10 return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver 15 polynucleotides encoding SECP to cells which have one or more genetic abnormalities with respect to the expression of SECP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are 20 described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nur. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver 25 polynucleotides encoding SECP to target cells which have one or more genetic abnormalities with respect to the expression of SECP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing SECP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has 30 been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under

the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding SECP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for SECP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of SECP-coding RNAs and the synthesis of high levels of SECP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of SECP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, 5 engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding SECP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, 10 corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by 15 any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding SECP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize 20 complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiester linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be 25 extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding SECP. Compounds 30 which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of

polynucleotide expression. Thus, in the treatment of disorders associated with increased SECP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding SECP may be therapeutically useful, and in the treatment of disorders associated with decreased SECP expression or activity, a compound which specifically promotes expression of the 5 polynucleotide encoding SECP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary 10 library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding SECP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted 15 biochemical system. Alterations in the expression of a polynucleotide encoding SECP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding SECP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one 20 or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a 25 human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691). 30 Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. 35 Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which 5 generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of SECP, antibodies to SECP, and mimetics, agonists, antagonists, or inhibitors of SECP.

10 The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. 15 These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. 20 Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

25 Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising SECP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, SECP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into 30 the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route

of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example SECP or fragments thereof, antibodies of SECP, and agonists, antagonists or inhibitors of SECP, which

- 5 ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large
- 10 therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.
- 15 The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy.
- 20 Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μ g to 100,000 μ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art.

- 25 Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polymucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind SECP may be used for the diagnosis

- 30 of disorders characterized by expression of SECP, or in assays to monitor patients being treated with SECP or agonists, antagonists, or inhibitors of SECP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for SECP include methods which utilize the antibody and a label to detect SECP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be

labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring SECP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of SECP expression. Normal or 5 standard values for SECP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to SECP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of SECP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation 10 between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding SECP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and 15 quantify gene expression in biopsied tissues in which expression of SECP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of SECP, and to monitor regulation of SECP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding SECP or closely related molecules may be used to identify nucleic acid sequences which encode SECP. The specificity of the probe, whether it is made 20 from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding SECP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% 25 sequence identity to any of the SECP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:11-20 or from genomic sequences including promoters, enhancers, and introns of the SECP gene.

Means for producing specific hybridization probes for DNAs encoding SECP include the cloning of polynucleotide sequences encoding SECP or SECP derivatives into vectors for the 30 production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding SECP may be used for the diagnosis of disorders associated with expression of SECP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal

- 5 hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an
- 10 autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis; anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema,
- 15 episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic
- 20 anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic
- 25 valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease,
- 30 aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and
- 35 other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary

ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss. The polynucleotide sequences encoding SECP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered SECP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding SECP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding SECP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding SECP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the

treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of SECP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a 5 fragment thereof, encoding SECP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the 10 presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several 15 days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ 20 preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding SECP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding SECP, 25 or a fragment of a polynucleotide complementary to the polynucleotide encoding SECP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences 30 encoding SECP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (FSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding SECP are used to amplify DNA using the

polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are 5 fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (iSNP), are capable of identifying polymorphisms by comparing the sequence of individual 10 overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing 15 errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of SECP include radiolabeling or biotinylation nucleotides, coamplification of a control nucleic acid, and interpolating results from 15 standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid 20 quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large 25 numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used 30 to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, SECP, fragments of SECP, or antibodies specific for SECP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to

generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number

5 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The

10 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention

15 may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000)

20 Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested

25 compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for

30 example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample

containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with 5 levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome 10 can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by 15 isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently 20 positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of 25 at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for SECP to quantify the levels of SECP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the 30 levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoza, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should

be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the

5 proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein

10 is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

15 In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

20 Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well

25 known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

30 In another embodiment of the invention, nucleic acid sequences encoding SECP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1

constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a 5 particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) 10 World Wide Web site. Correlation between the location of the gene encoding SECP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. 15 Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences 20 mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, SECP, its catalytic or immunogenic fragments, or 25 oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between SECP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds 30 having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with SECP, or fragments thereof, and washed. Bound SECP is then detected by methods well known in the art. Purified SECP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively,

non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding SECP specifically compete with a test compound for binding SECP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more 5 antigenic determinants with SECP.

In additional embodiments, the nucleotide sequences which encode SECP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

10 Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

15 The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/176,113, U.S. Ser. No. 60/177,733, U.S. Ser. No. 60/178,832, U.S. Ser. No. 60/179,774, and U.S. Ser. No. 60/186,792, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database 20 (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and 25 ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was 30 isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the

recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 5 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto 10 CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5a, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision 15 using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid 20 purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) *Anal. Biochem.* 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using 25 PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation 30 such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading 5 frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based 10 on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. 15 See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding 20 sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length 25 polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are 30 generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold

parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match 5 between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:11-20. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and 10 amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative secreted proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of 15 organisms (See Burge, C. and S. Karlin (1997) *J. Mol. Biol.* 268:78-94, and Burge, C. and S. Karlin (1998) *Curr. Opin. Struct. Biol.* 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA 20 sequences encode secreted proteins, the encoded polypeptides were analyzed by querying against PFAM models for secreted proteins. Potential secreted proteins were also identified by homology to Incyte cDNA sequences that had been annotated as secreted proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. 25 BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte 30 cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory

5 and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals

10 were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were

15 given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

20 **"Stretched" Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis

25 to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the

30 public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of SECP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:11-20 were compared with

sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:11-20 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available 5 from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map 10 position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation 15 hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression
20 Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related 25 molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

30
$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length(Seq. 1)}, \text{length(Seq. 2)} \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated

as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding SECP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding SECP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of SECP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction 5 mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the 10 alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

15 The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

20 The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels. fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were 25 religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

30 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified

using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

5 In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:11-20 are employed to screen cDNAs, genomic 10 DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - 32 P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston 15 MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

20 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and 25 compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned 30 technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may

contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may 5 comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a 10 fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

15 **Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M 20 dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and 25 incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and 30 resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses

primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

5 Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C
10 oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

15 Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

20 Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly
25 larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

30 Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a
35

resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the SECP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring SECP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of SECP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is

designed to prevent ribosomal binding to the SECP-encoding transcript.

XII. Expression of SECP

Expression and purification of SECP is achieved using bacterial or virus-based expression systems. For expression of SECP in bacteria, cDNA is subcloned into an appropriate vector containing

- 5 an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element.
- 10 Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express SECP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG).
- 15 Expression of SECP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding SECP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of
- 20 cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, SECP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from SECP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified SECP obtained by these methods can be used directly in the assays shown in

- 25 Examples XVI and XVII, where applicable.
- 30

XIII. Functional Assays

SECP function is assessed by expressing the sequences encoding SECP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice

include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the 5 apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in 10 expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. 15 (1994) Flow Cytometry, Oxford, New York NY.

The influence of SECP on gene expression can be assessed using highly purified populations of 20 cells transfected with sequences encoding SECP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression 25 of mRNA encoding SECP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of SECP Specific Antibodies

SECP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., 30 Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the SECP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well

described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to

5 increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-SECP activity by, for example, binding the peptide or SECP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring SECP Using Specific Antibodies

10 Naturally occurring or recombinant SECP is substantially purified by immunoaffinity chromatography using antibodies specific for SECP. An immunoaffinity column is constructed by covalently coupling anti-SECP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

15 Media containing SECP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of SECP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/SECP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and SECP is collected.

20 **XVI. Identification of Molecules Which Interact with SECP**

SECP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled SECP, washed, and any wells with labeled SECP complex are assayed. Data obtained using different concentrations of 25 SECP are used to calculate values for the number, affinity, and association of SECP with the candidate molecules.

Alternatively, molecules interacting with SECP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

30 SECP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of SECP Activity

An assay for growth stimulating or inhibiting activity of SECP measures the amount of DNA synthesis in Swiss mouse 3T3 cells (McKay, I. and Leigh, I., eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY). In this assay, varying amounts of SECP are added to quiescent 3T3 cultured cells in the presence of [³H]thymidine, a radioactive DNA precursor.

5 SECP for this assay can be obtained by recombinant means or from biochemical preparations. Incorporation of [³H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold SECP concentration range is indicative of growth modulating activity. One unit of activity per milliliter is defined as the 10 concentration of SECP producing a 50% response level, where 100% represents maximal incorporation of [³H]thymidine into acid-precipitable DNA.

Alternatively, an assay for SECP activity measures the stimulation or inhibition of neurotransmission in cultured cells. Cultured CHO fibroblasts are exposed to SECP. Following endocytic uptake of SECP, the cells are washed with fresh culture medium, and a whole cell voltage-15 clamped Xenopus myocyte is manipulated into contact with one of the fibroblasts in SECP-free medium. Membrane currents are recorded from the myocyte. Increased or decreased current relative to control values are indicative of neuromodulatory effects of SECP (Morimoto, T., et al. (1995) *Neuron* 15: 689-696).

20 Alternatively, the ADP-ribosylating activity of SECP is measured using the incorporation of radiolabeled ADP-ribose into proteins. For example, Barth, H. et al. (1998; *J. Biol. Chem.* 273:29506-11) incubated SECP with human platelet cytosol, containing about 50 µg protein, in 25 µL of reaction buffer (35 mM HEPES pH 7.5, 0.2 mM MgCl₂, 0.1 mM dithiothreitol, and 0.5 µM [*adenylate*-³²P]NAD for up to 10 minutes at 37 °C. Radiolabeled proteins were separated by 11% SDS-PAGE and quantified by phosphorimaging.

25 Alternatively, an assay for SECP activity measures the amount of SECP in secretory, membrane-bound organelles. Transfected cells as described above are harvested and lysed. The lysate is fractionated using methods known to those of skill in the art, for example, sucrose gradient ultracentrifugation. Such methods allow the isolation of subcellular components such as the Golgi apparatus, ER, small membrane-bound vesicles, and other secretory organelles. Immunoprecipitations 30 from fractionated and total cell lysates are performed using SECP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The concentration of SECP in secretory organelles relative to SECP in total cell lysate is proportional to the amount of SECP in transit through the secretory pathway.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention.

Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments.

5 Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	PolyPeptide SEQ ID NO:	Incyte Peptide ID	Polynucleotide SEQ ID NO:	Incyte Nucleotide ID
2508226	1	2508226CD1	11	2508226CB1
2880964	2	2880964CD1	12	2880964CB1
3122091	3	3122091CD1	13	3122091CB1
1633313	4	1633313CD1	14	1633313CB1
2323763	5	2323763CD1	15	2323763CB1
3759234	6	3759234CD1	16	3759234CB1
5583544	7	5583544CD1	17	5583544CB1
3459449	8	3459449CD1	18	3459449CB1
2544505	9	2544505CD1	19	2544505CB1
1568561	10	1568561CD1	20	1568561CB1

Table 2

PolyPeptide SEQ ID NO	Incyte PolyPeptide ID	GenBank ID Number	Probability Score	GenBank Homolog
1	2508226CD1	96635858	1.00E-176	lysophosphatidic acid phosphatase [Mus musculus]
2	2880564CD1	92648830	2.0E-19	dinitrogenase reductase activating glycohydrolase protein [Archaeoglobus fulgidus] (Klenk, H.P. et al. (1997) Nature 390:364-370)
3	3122091CD1	91789065	3.20E-15	putative oxidoreductase [Escherichia coli]
4	1633313CD1	92745735	9.40E-12	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase-I [Mus musculus] (Hennet, T. et al. (1998) J. Biol. Chem. 273:58-65)
5	2323763CD1	99657407	3.00E-11	chitinase [Vibrio cholerae]
6	3759234CD1	96094655	8.20E-106	similar to calmodulin; similar to P24044 (IPID:9115520)
		93413830	2.90E-19	Caldendrin [Rattus norvegicus] (Seidenbecher, C.I. et al. (1998) J. Biol. Chem. 273:21324-21331)
7	5583544CD1	9688297	3.50E-66	VDUP1 1,25-dihydroxyvitamin D-3 up-regulated (Homo sapiens) (Chen, K.S. and H.P. DeLuca (1994) Biochim. Biophys. Acta 1219,26-32)
8	3459449CD1	92677618	3.10E-83	NAD(P) (+) -arginine ADP-ribosyltransferase [Mus musculus]
10	1568561CD1	94322374	0	hephaestin [Mus musculus] (Vulpe, C.D., et al. (1999) Nat. Genet. 21 (2), 195-199)

Table 3

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences and Motifs	Analytical Methods and Databases
1	2508226CD1	428	T87 S104 T221 T320 S7 T114 S172 S210 S244 Y287		Signal peptide: M1-A32 Histidine acid phosphatase: R58-L65, H334-I347, H25-Y110, V53-K66, F120-F181, K324-A356, R50-P64 MAMMALIAN ACID PHOSPHATASE DM07177 P15309: G122-E428 P11117: G98-L399 HYDROLASE PD003234: S104-F181, T320-V374, V53-P64	BLAST_DOMO BLAST_PRODOME HMMER HMMER_PFAM BLIMPS_BLOCKS ProfileScan
2	2880964CD1	363	S67 S243 S38 S166 S206 S228 T259 Y116		Signal_cleavage: M1-A16 HYDROLASE PD094174: Y75-H182 HYDROLASE DINITROGENASE REDUCTASE PD008654:R22-Y282	SPScan BLAST_PRODOME
3	3122091CD1	373	T72 S187 S206 S210 T228 T291 S312 S364 T367	N248	FAD-dependent Pyridine nucleotide reductase signature PR00368A-E: H13-V35, L105-P114, V280-A287 PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES CLASS-I DM00071 P37596 1-241;V14-K246	BLIMPS-PRINTS BLAST-DOMO
4	1633313CD1	117	S15 S103		Signal_cleavage:M1-A26 Pyridine nucleotide-disulphide oxidoreductase PYR_redox:V14-L359 Signal peptide: M1-P49 UDP GlcNAc beta glycosyltransferase PD004190	SP SCAN HMMER-PFAM SPSCAN BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences and Motifs	Analytical Methods and Databases
5	2323763CD1	1049	S191 S966 S67 T136 S175 S191 S312 T426 S433 T434 T478 S527 T534 T574 T616 S623 T668 S708 S717 T729 S773 T778 S924 T974 S1003 S1005 T1015 S1043 S24 S115 S160 S270 T397 S543 T775 T804 S838 T901 T984	N207 N247 N395 N472 N487 N492 N525 N649 N742 N799 N890 N1029	Signal peptide: M1-A49 Transmembrane domain: Y932-C953 Chitinase A precursor PD036147	SPSCAN HMMER BLAST-PRODOM
6	3759234CD1	455	S39 T87 T103 S125 T144 S147 S158 S159 S386 S83 T241 S294 T354 T369 T380 T396 T404		EF hand: D286-E297, D322-E333 S100 Ca binding protein: I317-L337 Signal peptide: M1-C52 Transmembrane domain: L429-N447 EF hand: E277-L305, E313-K341 EF-hand calcium binding: D322-F334 S100/ICaBP type calcium binding: N310-G346 S100 Ca binding protein: L305-V359 Calcium binding repeat protein PD000012: R274-L342	MOTIFS MOTIFS SPSCAN HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-BLOCKS PROFILESCAN BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences and Motifs	Analytical Methods and Databases
7	5583544CD1	407	S74 T118 T150 T196 S389 S76 S83 S126 S187 S211 T291	N209	Signal peptide: M1-A46 Similar to dihydroxyvitamin D3 induced protein PD004148; E21-A229	SPSCAN BLAST-PRODOM
8	3459449CD1	240	T35 T75 S165 S176 S185 S212	N102	Arginine ADP-ribosyltransferase: PR00970 A:G30-A51 B:E58-W76 C:F89-S103 D:F129-L145 E:V177-A192	BLIMPS-PRINTS
9	2544505CD1	97	S59 S76 T33 S35 S51 S54 T68 S71		NAD:arginine ADP-ribosyltransferase BL01291 A:L29-E58 B:S68-R81 C:P86-T101 D:L11-L144 E:V159-F168 Signal_Peptide:M1-A22 Signal_cleavage:M1-G14 NAD:arginine ADP-ribosyltransferase: ART:L3-G237 Glycosaminoglycan: S236-G239 ADP-RIBOSYLTRANSFERASE: PD004385:L6-L235 RIBOSYLTRANSFERASE: DM02464 P52961 1-326:L27-L235 Signal_Peptide: M1-A26 Prostanoid EP3 receptor signature: A86-G97	BLIMPS-BLOCKS HMMER SPScan HMMER-PFAM BLAST-PRODOM BLAST-DOMO

Table 3 (cont.)

SEQ NO:	Incyte ID	Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences and Motifs	Analytical Methods and Databases
10	1568561CD1	1161	S88 T169 T177 S190 S234 T264 S414 T443 T531 S544 T586 T654 S705 S763 S774 T804 S887 S918 T960 S1067 S1148 S67 T75 T208 S268 T317 T379 T405 S416 S424 T519 S620 T727 T796 S964 T1082 S1153 S1158 Y89 Y144 Y501 Y520	N167 N239 N591 N717 N761 N832 N876 N934	Transmembrane domain: V1108-A1128 Multicopper oxidase domain: L231-C369; L583-C721; L926-R1068 Multicopper oxidases protein signature: G460-N477; L1102-F1012; N1041-M1059 Multicopper oxidases 1 signatures: D324-Q384; G676-R729; D1023-A1085 Multicopper oxidases 2 signatures: F329-K418; F681-Q730; F1028-T1079 FACTOR PRECURSOR GLYCOPROTEIN PLASMA REPEAT SIGNAL COAGULATION BLOOD VIII CALCIUM P002090; T28-D224; K381-K575; Q725-D921 FERROXIDASE REPEAT D00956 P00450190-336; L99-SJ49; L456-G696; L810-M1047 Multicopper Oxidase 1 motif: G344-Y364; G696-Y716; G1043-F1063 Multicopper Oxidase 2 motif: H1048-M1059	BLIMPS-BLOCKS HMMER-PFM HMMER-PFM	

Table 4

Nucleotide SEQ ID NO	Incyte Nucleotide ID	Sequence length	Selected fragment(s)	Sequence fragments	5' position	3' position
11	2508226CB1	1806	135-564, 1-60, 1047-1200	60137216V1 6294749H1 (BMARUNA01) 2313881T6 (INGANNOT01) 2509226F6 (CONUTUT01) 6015394H1 (PITUNOT06) 60137216D1	1415 588 1182 1 889 291	1806 905 1767 338 1580 794
12	2880964CB1	1271		1306721F6 (PLACNOT02) 2564331H1 (ADRETUT01) 4767612H1 (BRATNOT02) 60205253U1	226 1009 956 512	828 1271 1216 1001
13	3122091CB1	1402	1-113	4407267H1 (PROSDIT01) 3988634H1 (LUNGNON03) 6361840H1 (CONFNOT09) 6839128H1 (BRSTNON02) 3122091F6 (LNODNOT05)	1 1 29 753 329	266 294 704 1402 759
14	16333313CB1	1156	414-604, 1-254	6836430H1 (BRSTNON02) 1633313F6 (COLNNOT19) 6831516J1 (SINTNOR01) 1801159H1 (SINTNOT13) 6576579H1 (COLHTUS02)	632 1 563 931 356	1267 431 1151 1156 985

Table 4

15	2323763CB1	3908	2421-2457, 353-749, 1281-1626	6706202H1 (HEADDIR01) 5840312H1 (BRAENOT04) 6782259H1 (OVARIEDR01) 6984527H1 (braifer05) 60209615U1	3261 2552 2014 1403 3076	3853 2843 2580 2007 3703
16	3759234CB1	1368	80-805, 1254-1368	7071368H1 (BRAUDR02) 7091963H1 (BRAUDR03) 6931827H1 (SINTMTR02) 6917482H1 (PLACFER06)	575 1 1831 2567	1170 466 2536 3178
17	5583544CB1	2827	1286-2051, 1-453	95523837.v113.9s 3. nt. edit 6618758H1 (BRAIDR02) 6830118H1 (SINTNOR01) 6758583J1 (HEAONOR01) 6830118J1 (SINTNOR01) 6542607H1 (LNODN0N02)	984 1 986 984 1	3908 1086 1290 1086 1368
18	3459449CB1	1113	1-149, 497-535, 390-461	6299221H1 (UTREDIT07) 2231153T6 (PROSNOT16) 6573737H1 (COLHTUS02) 5824487H1 (MCLRUNT03)	1598 1 713 318 2261	2231 564 1254 1002 2827
19	2544505CB1	569	1-569	3663102F6 (PANCNOT16) 70475093V1 70475907V1 3597042F6 (FIBPN0T01)	1 444 574 12	503 1006 1113 569

Table 4

20	1568561CB1	4427	3652-3783 1152-1736, 1-21, 1898-2478, 477-1021, 2965-3084	7056696H1 (BRALNON02) 20B3063F6 (UTRSNOT08) 2708849F6 (PONSAZT01) 70879535V1 70743888V1 2681403F6 (SINIUCT01) 70741272V1 70744146V1 70879568V1 1500127F6 (SINTBST01) 6356852H1 (LUNGDIS03) 70880086V1 6755427H1 (SINTFER02) 6755427J1 (SINTFER02)	3843 1002 1 2377 3082 1570 1824 551 3599 3707 505 1182 2708 2117 2857	4427 1529 562 2901 3678 2101 2276 1024 4193 4295 795 1695 3443 2857
----	------------	------	--	--	---	--

Table 4 (cont.)

Nucleotide SEQ ID NO	Incyte Nucleotide ID	Sequence length	Selected fragment (s)	Sequence fragments	5' position	3' position
11	2508226CB1	1806	135-564, 1-60, 1047-1200	60137216V1 6294749H1 (BMARUNA01) 2313881T6 (INGANNNOT01) 2508226F6 (CONUTUT01) 6035394H1 (PITUNOT06) 60137216D1	1415 588 1182 1 889	1806 905 1767 338 1580
12	2880964CB1	1271		1306721F6 (PLACNOT02) 2564331H1 (ADREUTU01) 4767612H1 (BRAINTNOT02) 60205253U1	226 1009 956 512	828 1271 1216 1001
13	3122091CB1	1402	1-113	4407267H1 (PROSDIT01) 3988634H1 (LUNGNON03) 6361840H1 (CONFNOT09) 6839128H1 (BRSTHNON02)	1 1 29 704	266 294 704 1402
14	16331313CB1	1156	414-604, 1-254	3122091F6 (LNODNOT05) 6836430H1 (BRSTHNON02) 1633313F6 (COLNNOT19) 6831516J1 (SINTNOR01) 1803159H1 (SINTNOT13) 6576579H1 (COLHTUS02)	329 632 1 563 931 356	759 1267 431 1151 1156 985

Table 4 (cont.)

15	2323763CB1	3908	2421-2457, 353-749, 1281-1626	6706202H1 (HEAADIR01) 5840312H1 (BRAENOT04) 6782259H1 (OVARDIR01) 6984527H1 (braifer05) 60209615U1	3261 2552 2014 1403 3076	3853 2843 2580 2007 3703
16	3759234CB1	1368	80-805, 1254-1368	7091963H1 (BRAUTDR01) 6931827H1 (SINTTMR02) 6917482H1 (PLACFER06) 3759234H1 (BRAHDIT03)	575 1 1831 2567	1170 466 2536 3178
17	5583544CB1	2827	1286-2051, 1-453	4244667F8 (BRABDIT01) 95523837_v113_gs_3_nt.edit 6618758H1 (BRAITDR02) 6830118H1 (SINTNOR01)	984 1 1598 1	1086 1368 2231 564
18	3459449CB1	1113	1-149, 497-535, 390-461	6758583J1 (HEAONOR01) 6830118J1 (SINTNOR01) 6542607H1 (LNODNON02) 6299221H1 (UTREDIT07) 2231153T6 (PROSNOT16) 6573737H1 (COLHTUS02) 5824487H1 (MCLRUNIT03)	713 318 2261 1204 2193 1239 1656	1254 1002 2827 1478 2809 1656 2275
19	2544505CB1	569	1-569	3663102F6 (PANCNOT16) 70475093V1 70475907V1 3597042F6 (FIBPNOT01)	1 444 574 12	503 1006 1113 569

Table 4 (cont.)

20	1568561CB1	4427	3652-3783, 1152-1736, 1-21, 1898-2478, 477-1021, 2965-3084	7056696H1 (BRAUNON02) 2083063F6 (UTRENC0708) 2708849F6 (PONSAZT01) 70879535V1 70743888V1 2681403F6 (SINNUCT01) 70741272V1 70744146V1 70879568V1 1500127F6 (SINTBST01) 6356852H1 (LUNGDIS03) 70880086V1 675542H1 (SINTFER02) 6755427J1 (SINTFER02)	3843 1002 1 2377 3082 1570 1824 551 3599 3707 505 1182 2708 2117	4427 1529 562 2901 3678 2101 2276 1024 4193 4295 795 1695 3443 2857
----	------------	------	---	--	---	--

Table 5

Nucleotide SEQ ID NO:	Incite Project ID	Representative Library
11	2503226CB1	EOSIHETO2
12	2880964CB1	BRAYDINO3
13	3122091CB1	LIVRNNO8
14	1633113CB1	COLANNOT19
15	2323763CB1	HEAANOT01
16	3759234CB1	BRAHDITO3
17	5583544CB1	MCLRUNTO1
18	3459419CB1	HNT2UNNO3
19	2544505CB1	FIBPNOT01
20	1568561CB1	SINTBST01

Table 6

Library	Vector	Library Description
BRAHDIT03	PINCY	Library was constructed using RNA isolated from diseased hippocampus brain tissue. Patient history included Alzheimer's disease.
BRAYDIN03	PINCY	This normalized brain tissue library was constructed from 6.7 million independent clones from the BRAYDIT01 tissue library. Starting RNA was made from RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 -hours/round) reannealing hybridization was used. The library was linearized and recircularized to select for insert containing clones.
COLNN019	PINCY	Library was constructed using RNA isolated from the cecal tissue of an 18-year-old Caucasian female. The cecal tissue, along with the appendix and ileum tissue, were removed during bowel anastomosis. Pathology indicated Crohn's disease of the ileum, involving 15 cm of the small bowel.
EOSTHET02	PBLUESCRIPT	Library was constructed using RNA isolated from peripheral blood cells apheresed from a 48-year-old Caucasian male. Patient history included hypereosinophilia. The cell population was determined to be greater than 77% eosinophils by Wright's staining.
FIBBPNOT01	PINCY	Library was constructed using RNA isolated from fibroblasts of the prostate stroma removed from a male fetus, who died after 26 weeks' gestation.

Table 6 (cont.)

Library	Vector	Library Description
HEANOT01	PINCY	Library was constructed using RNA isolated from right coronary and right circumflex coronary artery tissue removed from the explanted heart of a 46-year-old Caucasian male during a heart transplantation. Patient history included myocardial infarction from total occlusion of the left anterior descending coronary artery, atherosclerotic coronary artery disease, hyperlipidemia, myocardial ischemia, dilated cardiomyopathy, left ventricular dysfunction, and tobacco abuse. Previous surgeries included cardiac catheterization. Family history included atherosclerotic coronary artery disease.
HNT2UNN03	PSPORT1	This normalized NT2 cell line library was constructed from independent clones from an untreated NT2 cell line library. Starting RNA was made from the NT2 cell line derived from a human teratocarcinoma, which exhibited properties characteristic of a committed neuronal precursor at an early stage of development. The cells were untreated. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996) 791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LIVRN08	PINCY	This normalized library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was made from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16-week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996) 791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 6 (cont.)

Library	Vector	Library Description
MCLRUNT01	pINCY	The library was constructed using RNA isolated from untreated peripheral blood mononuclear cell tissue obtained from buffy coat and removed from a 60-year-old male.
SINTBST01	pINCY	Library was constructed using RNA isolated from ileum tissue obtained from an 18-year-old Caucasian female during bowel anastomosis. Pathology indicated Crohn's disease of the ileum, involving 15 cm of the small bowel. Family history included cerebrovascular disease and atherosclerotic coronary artery disease.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25: 3389-3402.	<i>ESTs</i> : Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
92 FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, ifasta, fastx, ifastx, and search.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad. Sci.</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:481-489.	<i>ESTs</i> : fasta E value= 1.06E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fasta E value= 1.0E-8 or less <i>Full Length sequences</i> : fasta score=100 or greater
BLIMPS	A BLocks IMproved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff. <i>Nucl. Acid Res.</i> , 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) <i>Methods Enzymol.</i> 266:88-105; and Altwood, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Sonnhammer, E.L.L. et al. (1998) <i>Nucleic Acids Res.</i> 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score≥GCG-specified "HICRF" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-183; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202...	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. <i>supra</i> ; Wisconsin Package Program Manual, Version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,
 - b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,
 - c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and
 - 10 d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-10.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 20 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:11-20.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 30 9. A method for producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
 - 35 b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:
 - 5 a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,
 - b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,
 - c) a polynucleotide sequence complementary to a),
 - d) a polynucleotide sequence complementary to b), and
 - 10 e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
- 15 13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
 - a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
 - 20 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
- 25 15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
 - a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
 - 30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

18. A method for treating a disease or condition associated with decreased expression of 5 functional SECP, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

10 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

15 21. A method for treating a disease or condition associated with decreased expression of functional SECP, comprising administering to a patient in need of such treatment a composition of claim 20.

20 22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting antagonist activity in the sample.

25 23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional SECP, comprising administering to a patient in need of such treatment a composition of claim 23.

30 25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- 5 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in 10 the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method 15 comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of 20 the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at 25 least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

<110> INCYTE GENOMICS, INC.
 BANDMAN, Olga
 PATTERSON, Chandra
 WALIA, Narinder K.
 NGUYEN, Dannie B.
 YUE, Henry
 KHAN, Farrah A.
 TANG, Y. Tom
 BAUGHN, Mariah R.
 LU, Dyung Aina M.
 YANG, Junming
 BURFORD, Neil
 AU-YOUNG, Janice
 REDDY, Roopa

<120> SECRETED PROTEINS

<130> PI-0010 PCT

<140> To Be Assigned
 <141> Herewith

<150> 60/176,113; 60/177,733; 60/178,832; 60/179,774; 60/186,792
 <151> 2000-01-14; 2000-01-21; 2000-01-28; 2000-02-02; 2000-03-03

<160> 20

<170> PERL Program

<210> 1

<211> 428

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2508226CD1

<400> 1
 Met Ile Thr Gly Val Phe Ser Met Arg Leu Trp Thr Pro Val Gly
 1 5 10 15
 Val Leu Thr Ser Leu Ala Tyr Cys Leu His Gln Arg Arg Val Ala
 20 25 30
 Leu Ala Glu Leu Gln Glu Ala Asp Gly Gln Cys Pro Val Asp Arg
 35 40 45
 Ser Leu Leu Lys Leu Lys Met Val Gln Val Val Phe Arg His Gly
 50 55 60
 Ala Arg Ser Pro Leu Lys Pro Leu Pro Leu Glu Glu Gln Val Glu
 65 70 75
 Trp Asn Pro Gln Leu Leu Glu Val Pro Pro Gln Thr Gln Phe Asp
 80 85 90
 Tyr Thr Val Thr Asn Leu Ala Gly Gly Pro Lys Pro Tyr Ser Pro
 95 100 105
 Tyr Asp Ser Gln Tyr His Glu Thr Thr Leu Lys Gly Gly Met Phe
 110 115 120
 Ala Gly Gln Leu Thr Lys Val Gly Met Gln Gln Met Phe Ala Leu
 125 130 135
 Gly Glu Arg Leu Arg Lys Asn Tyr Val Glu Asp Ile Pro Phe Leu
 140 145 150
 Ser Pro Thr Phe Asn Pro Gln Glu Val Phe Ile Arg Ser Thr Asn
 155 160 165
 Ile Phe Arg Asn Leu Glu Ser Thr Arg Cys Leu Leu Ala Gly Leu
 170 175 180
 Phe Gln Cys Gln Lys Glu Gly Pro Ile Ile Ile His Thr Asp Glu
 185 190 195
 Ala Asp Ser Glu Val Leu Tyr Pro Asn Tyr Gln Ser Cys Trp Ser
 200 205 210

Leu Arg Gln Arg Thr Arg Gly Arg Arg Gln Thr Ala Ser Leu Gln
 215 220 225
 Pro Gly Ile Ser Glu Asp Leu Lys Lys Val Lys Asp Arg Met Gly
 230 235 240
 Ile Asp Ser Ser Asp Lys Val Asp Phe Phe Ile Leu Leu Asp Asn
 245 250 255
 Val Ala Ala Glu Gln Ala His Asn Leu Pro Ser Cys Pro Met Leu
 260 265 270
 Lys Arg Phe Ala Arg Met Ile Glu Gln Arg Ala Val Asp Thr Ser
 275 280 285
 Leu Tyr Ile Leu Pro Lys Glu Asp Arg Glu Ser Leu Gln Met Ala
 290 295 300
 Val Gly Pro Phe Leu His Ile Leu Glu Ser Asn Leu Leu Lys Ala
 305 310 315
 Met Asp Ser Ala Thr Ala Pro Asp Lys Ile Arg Lys Leu Tyr Leu
 320 325 330
 Tyr Ala Ala His Asp Val Thr Phe Ile Pro Leu Leu Met Thr Leu
 335 340 345
 Gly Ile Phe Asp His Lys Trp Pro Pro Phe Ala Val Asp Leu Thr
 350 355 360
 Met Glu Leu Tyr Gln His Leu Glu Ser Lys Glu Trp Phe Val Gln
 365 370 375
 Leu Tyr Tyr His Gly Lys Glu Gln Val Pro Arg Gly Cys Pro Asp
 380 385 390
 Gly Leu Cys Pro Leu Asp Met Phe Leu Asn Ala Met Ser Val Tyr
 395 400 405
 Thr Leu Ser Pro Glu Lys Tyr His Ala Leu Cys Ser Gln Thr Gln
 410 415 420
 Val Met Glu Val Gly Asn Glu Glu
 425

<210> 2
 <211> 363
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2880964CD1

<400> 2
 Met Ala Ala Ala Ala Met Ala Ala Ala Gly Gly Gly Ala Gly
 1 5 10 15
 Ala Ala Arg Ser Leu Ser Arg Phe Arg Gly Cys Leu Ala Gly Ala
 20 25 30
 Leu Leu Gly Asp Cys Val Gly Ser Phe Tyr Glu Ala His Asp Thr
 35 40 45
 Val Asp Leu Thr Ser Val Leu Arg His Val Gln Ser Leu Glu Pro
 50 55 60
 Asp Pro Gly Thr Pro Gly Ser Glu Arg Thr Glu Ala Leu Tyr Tyr
 65 70 75
 Thr Asp Asp Thr Ala Met Ala Arg Ala Leu Val Gln Ser Leu Leu
 80 85 90
 Ala Lys Glu Ala Phe Asp Glu Val Asp Met Ala His Arg Phe Ala
 95 100 105
 Gln Glu Tyr Lys Lys Asp Pro Asp Arg Gly Tyr Gly Ala Gly Val
 110 115 120
 Val Thr Val Phe Lys Lys Leu Leu Asn Pro Lys Cys Arg Asp Val
 125 130 135
 Phe Glu Pro Ala Arg Ala Gln Phe Asn Gly Lys Gly Ser Tyr Gly
 140 145 150
 Asn Gly Gly Ala Met Arg Val Ala Gly Ile Ser Leu Ala Tyr Ser
 155 160 165
 Ser Val Gln Asp Val Gln Lys Phe Ala Arg Leu Ser Ala Gln Leu
 170 175 180
 Thr His Ala Ser Ser Leu Gly Tyr Asn Gly Ala Ile Leu Gln Ala
 185 190 195

Leu Ala Val His Leu Ala Leu Gln Gly Glu Ser Ser Ser Glu His
 200 205 210
 Phe Leu Lys Gln Leu Leu Gly His Met Glu Asp Leu Glu Gly Asp
 215 220 225
 Ala Gln Ser Val Leu Asp Ala Arg Glu Leu Gly Met Glu Glu Arg
 230 235 240
 Pro Tyr Ser Ser Arg Leu Lys Lys Ile Gly Glu Leu Leu Asp Gln
 245 250 255
 Ala Ser Val Thr Arg Glu Glu Val Val Ser Glu Leu Gly Asn Gly
 260 265 270
 Ile Ala Ala Phe Glu Ser Val Pro Thr Ala Ile Tyr Cys Phe Leu
 275 280 285
 Arg Cys Met Glu Pro Asp Pro Glu Ile Pro Ser Ala Phe Asn Ser
 290 295 300
 Leu Gln Arg Thr Leu Ile Tyr Ser Ile Ser Leu Gly Gly Asp Thr
 305 310 315
 Asp Thr Ile Ala Thr Met Ala Gly Ala Ile Ala Gly Ala Tyr Tyr
 320 325 330
 Gly Met Asp Gln Val Pro Glu Ser Trp Gln Gln Ser Cys Glu Gly
 335 340 345
 Tyr Glu Glu Thr Asp Ile Leu Ala Gln Ser Leu His Arg Val Phe
 350 355 360
 Gln Lys Ser

<210> 3
 <211> 373
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3122091CD1

<400> 3
 Met Gly Ser Gln Val Ser Val Glu Ser Gly Ala Leu His Val Val
 1 5 10 15
 Ile Val Gly Gly Gly Phe Gly Gly Ile Ala Ala Ala Ser Gln Leu
 20 25 30
 Gln Ala Leu Asn Val Pro Phe Met Leu Val Asp Met Lys Asp Ser
 35 40 45
 Phe His His Asn Val Ala Ala Leu Arg Ala Ser Val Glu Thr Gly
 50 55 60
 Phe Ala Lys Lys Thr Phe Ile Ser Tyr Ser Val Thr Phe Lys Asp
 65 70 75
 Asn Phe Arg Gln Gly Leu Val Val Gly Ile Asp Leu Lys Asn Gln
 80 85 90
 Met Val Leu Leu Gln Gly Gly Glu Ala Leu Pro Phe Ser His Leu
 95 100 105
 Ile Leu Ala Thr Gly Ser Thr Gly Pro Phe Pro Gly Lys Phe Asn
 110 115 120
 Glu Val Ser Ser Gln Gln Ala Ala Ile Gln Ala Tyr Glu Asp Met
 125 130 135
 Val Arg Gln Val Gln Arg Ser Arg Phe Ile Val Val Val Gly Gly
 140 145 150
 Gly Ser Ala Gly Val Glu Met Ala Ala Glu Ile Lys Thr Glu Tyr
 155 160 165
 Pro Glu Lys Glu Val Thr Leu Ile His Ser Gln Val Ala Leu Ala
 170 175 180
 Asp Lys Glu Leu Leu Pro Ser Val Arg Gln Glu Val Lys Glu Ile
 185 190 195
 Leu Leu Arg Lys Gly Val Gln Leu Leu Leu Ser Glu Arg Val Ser
 200 205 210
 Asn Leu Glu Glu Leu Pro Leu Asn Glu Tyr Arg Glu Tyr Ile Lys
 215 220 225
 Val Gln Thr Asp Lys Gly Thr Glu Val Ala Thr Asn Leu Val Ile
 230 235 240

Leu Cys Thr Gly Ile Lys Ile Asn Ser Ser Ala Tyr Arg Lys Ala
 245 250 255
 Phe Glu Ser Arg Leu Ala Ser Ser Gly Ala Leu Arg Val Asn Glu
 260 265 270
 His Leu Gln Val Glu Gly His Ser Asn Val Tyr Ala Ile Gly Asp
 275 280 285
 Cys Ala Asp Val Arg Thr Pro Lys Met Ala Tyr Leu Ala Gly Leu
 290 295 300
 His Ala Asn Ile Ala Val Ala Asn Ile Val Asn Ser Val Lys Gln
 305 310 315
 Arg Pro Leu Gln Ala Tyr Lys Pro Gly Ala Leu Thr Phe Leu Leu
 320 325 330
 Ser Met Gly Arg Asn Asp Gly Val Gly Gln Ile Ser Gly Phe Tyr
 335 340 345
 Val Gly Arg Leu Met Val Arg Leu Thr Lys Ser Arg Asp Leu Phe
 350 355 360
 Val Ser Thr Ser Trp Lys Thr Met Arg Gln Ser Pro Pro
 365 370

<210> 4
 <211> 117
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1633313CD1

<400> 4
 Met Pro Leu Arg Lys Pro Gly Gly Pro Phe Tyr Val Pro Glu Ser
 1 5 10 15
 Phe Phe Glu Gly Gly Tyr Pro Ala Tyr Ala Ser Gly Gly Gly Tyr
 20 25 30
 Val Ile Ala Gly Arg Leu Ala Pro Trp Leu Leu Arg Ala Ala Ala
 35 40 45
 Arg Val Ala Pro Phe Pro Phe Glu Asp Val Tyr Thr Gly Leu Cys
 50 55 60
 Ile Arg Ala Leu Gly Leu Val Pro Gln Ala His Pro Gly Phe Leu
 65 70 75
 Thr Ala Trp Pro Ala Asp Arg Thr Ala Asp His Cys Ala Phe Arg
 80 85 90
 Asn Leu Leu Leu Val Arg Pro Leu Gly Pro Gln Ala Ser Ile Arg
 95 100 105
 Leu Trp Lys Gln Leu Gln Asp Pro Arg Leu Gln Cys
 110 115

<210> 5
 <211> 1049
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2323763CD1

<400> 5
 Met Glu Lys Arg Leu Gly Val Lys Pro Asn Pro Ala Ser Trp Ile
 1 5 10 15
 Leu Ser Gly Tyr Tyr Trp Gln Thr Ser Ala Lys Trp Leu Arg Ser
 20 25 30
 Leu Tyr Leu Phe Tyr Thr Cys Phe Cys Phe Ser Val Leu Trp Leu
 35 40 45
 Ser Thr Asp Ala Ser Glu Ser Arg Cys Gln Gln Gly Lys Thr Gln
 50 55 60
 Phe Gly Val Gly Leu Arg Ser Gly Glu Asn His Leu Trp Leu
 65 70 75
 Leu Glu Gly Thr Pro Ser Leu Gln Ser Cys Trp Ala Ala Cys Cys

80	85	90
Gln Asp Ser Ala Cys His Val Phe Trp Trp	Leu Glu Gly Met Cys	
95	100	105
Ile Gln Ala Asp Cys Ser Arg Pro Gln Ser	Cys Arg Ala Phe Arg	
110	115	120
Thr His Ser Ser Asn Ser Met Leu Val Phe	Leu Lys Lys Phe Gln	
125	130	135
Thr Ala Asp Asp Leu Gly Phe Leu Pro Glu	Asp Asp Val Pro His	
140	145	150
Leu Leu Gly Leu Gly Trp Asn Trp Ala Ser	Trp Arg Gln Ser Pro	
155	160	165
Pro Arg Ala Ala Leu Arg Pro Ala Val Ser	Ser Ser Asp Gln Gln	
170	175	180
Ser Leu Ile Arg Lys Leu Gln Lys Arg Gly	Ser Pro Ser Asp Val	
185	190	195
Val Thr Pro Ile Val Thr Gln His Ser Lys	Val Asn Asp Ser Asn	
200	205	210
Glu Leu Gly Gly Leu Thr Thr Ser Gly Ser	Ala Glu Val His Lys	
215	220	225
Ala Ile Thr Ile Ser Ser Pro Leu Thr Thr	Asp Leu Thr Ala Glu	
230	235	240
Leu Ser Gly Gly Pro Lys Asn Val Ser Val	Gln Pro Glu Ile Ser	
245	250	255
Glu Gly Leu Ala Thr Thr Pro Ser Thr Gln	Gln Val Lys Ser Ser	
260	265	270
Glu Lys Thr Gln Ile Ala Val Pro Gln Pro	Val Ala Pro Ser Tyr	
275	280	285
Ser Tyr Ala Thr Pro Thr Pro Gln Ala Ser	Phe Gln Ser Thr Ser	
290	295	300
Ala Pro Tyr Pro Val Ile Lys Glu Leu Val	Val Ser Ala Gly Glu	
305	310	315
Ser Val Gln Ile Thr Leu Pro Lys Asn Glu	Val Gln Leu Asn Ala	
320	325	330
Tyr Val Leu Gln Glu Pro Pro Lys Gly Glu	Thr Tyr Thr Tyr Asp	
335	340	345
Trp Gln Leu Ile Thr His Pro Arg Asp Tyr	Ser Gly Glu Met Glu	
350	355	360
Gly Lys His Ser Gln Ile Leu Lys Leu Ser	Lys Leu Thr Pro Gly	
365	370	375
Leu Tyr Glu Phe Lys Val Ile Val Glu Gly	Gln Asn Ala His Gly	
380	385	390
Glu Gly Tyr Val Asn Val Thr Val Lys Pro	Glu Pro Arg Lys Asn	
395	400	405
Arg Pro Pro Ile Ala Ile Val Ser Pro Gln	Phe Gln Glu Ile Ser	
410	415	420
Leu Pro Thr Thr Ser Thr Val Ile Asp Gly	Ser Gln Ser Thr Asp	
425	430	435
Asp Asp Lys Ile Val Gln Tyr His Trp Glu	Glu Leu Lys Gly Pro	
440	445	450
Leu Arg Glu Glu Lys Ile Ser Glu Asp Thr	Ala Ile Leu Lys Leu	
455	460	465
Ser Lys Leu Val Pro Gly Asn Tyr Thr Phe	Ser Leu Thr Val Val	
470	475	480
Asp Ser Asp Gly Ala Thr Asn Ser Thr Ala	Asn Leu Thr Val	
485	490	495
Asn Lys Ala Val Asp Tyr Pro Pro Val Ala	Asn Ala Gly Pro Asn	
500	505	510
Gln Val Ile Thr Leu Pro Gln Asn Ser Ile	Thr Leu Phe Gly Asn	
515	520	525
Gln Ser Thr Asp Asp His Gly Ile Thr Ser	Tyr Glu Trp Ser Leu	
530	535	540
Ser Pro Ser Ser Lys Gly Lys Val Val Glu	Met Gln Gly Val Arg	
545	550	555
Thr Pro Thr Leu Gln Leu Ser Ala Met Gln	Glu Gly Asp Tyr Thr	
560	565	570
Tyr Gln Leu Thr Val Thr Asp Thr Ile Gly	Gln Gln Ala Thr Ala	
575	580	585

Gln Val Thr Val Ile Val Gln Pro Glu Asn Asn Lys Pro Pro Gln
 590 595 600
 Ala Asp Ala Gly Pro Asp Lys Glu Leu Thr Leu Pro Val Asp Ser
 605 610 615
 Thr Thr Leu Asp Gly Ser Lys Ser Ser Asp Asp Gln Lys Ile Ile
 620 625 630
 Ser Tyr Leu Trp Glu Lys Thr Gln Gly Pro Asp Gly Val Gln Leu
 635 640 645
 Glu Asn Ala Asn Ser Ser Val Ala Thr Val Thr Gly Leu Gln Val
 650 655 660
 Gly Thr Tyr Val Phe Thr Leu Thr Val Lys Asp Glu Arg Asn Leu
 665 670 675
 Gln Ser Gln Ser Ser Val Asn Val Ile Val Lys Glu Glu Ile Asn
 680 685 690
 Lys Pro Pro Ile Ala Lys Ile Thr Gly Asn Val Val Ile Thr Leu
 695 700 705
 Pro Thr Ser Thr Ala Glu Leu Asp Gly Ser Lys Ser Ser Asp Asp
 710 715 720
 Lys Gly Ile Val Ser Tyr Leu Trp Thr Arg Asp Glu Gly Ser Pro
 725 730 735
 Ala Ala Gly Glu Val Leu Asn His Ser Asp His His Pro Ile Leu
 740 745 750
 Phe Leu Ser Asn Leu Val Glu Gly Thr Tyr Thr Phe His Leu Lys
 755 760 765
 Val Thr Asp Ala Lys Gly Glu Ser Asp Thr Asp Arg Thr Thr Val
 770 775 780
 Glu Val Lys Pro Asp Pro Arg Lys Asn Asn Leu Val Glu Ile Ile
 785 790 795
 Leu Asp Ile Asn Val Ser Gln Leu Thr Glu Arg Leu Lys Gly Met
 800 805 810
 Phe Ile Arg Gln Ile Gly Val Leu Leu Gly Val Leu Asp Ser Asp
 815 820 825
 Ile Ile Val Gln Lys Ile Gln Pro Tyr Thr Glu Gln Ser Thr Lys
 830 835 840
 Met Val Phe Phe Val Gln Asn Glu Pro Pro His Gln Ile Phe Lys
 845 850 855
 Gly His Glu Val Ala Ala Met Leu Lys Ser Glu Leu Arg Lys Gln
 860 865 870
 Lys Ala Asp Phe Leu Ile Phe Arg Ala Leu Glu Val Asn Thr Val
 875 880 885
 Thr Cys Gln Leu Asn Cys Ser Asp His Gly His Cys Asp Ser Phe
 890 895 900
 Thr Lys Arg Cys Ile Cys Asp Pro Phe Trp Met Glu Asn Phe Ile
 905 910 915
 Lys Val Gln Leu Arg Asp Gly Asp Ser Asn Cys Glu Trp Ser Val
 920 925 930
 Leu Tyr Val Ile Ile Ala Thr Phe Val Ile Val Val Ala Leu Gly
 935 940 945
 Ile Leu Ser Trp Thr Val Ile Cys Cys Cys Lys Arg Gln Lys Gly
 950 955 960
 Lys Pro Lys Arg Lys Ser Lys Tyr Lys Ile Leu Asp Ala Thr Asp
 965 970 975
 Gln Glu Ser Leu Glu Leu Lys Pro Thr Ser Arg Ala Gly Ile Lys
 980 985 990
 Gln Lys Gly Leu Leu Leu Ser Ser Ser Leu Met His Ser Glu Ser
 995 1000 1005
 Glu Leu Asp Ser Asp Asp Ala Ile Phe Thr Trp Pro Asp Arg Glu
 1010 1015 1020
 Lys Gly Lys Leu Leu His Gly Gln Asn Gly Ser Val Pro Asn Gly
 1025 1030 1035
 Gln Thr Pro Leu Lys Ala Arg Ser Pro Arg Glu Glu Ile Leu
 1040 1045

<210> 6
 <211> 455
 <212> PRT
 <213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3759234CD1

<400> 6

Met	Tyr	Arg	Gly	Ile	Tyr	Thr	Val	Pro	Asn	Leu	Leu	Ser	Glu	Gln
1				5					10				15	
Arg	Pro	Val	Asp	Ile	Pro	Glu	Asp	Glu	Leu	Glu	Gly	Glu	Cys	Pro
					20				25				30	
Pro	Gly	Ser	Pro	Pro	Arg	Arg	Pro	Ser	Tyr	Leu	Cys	Ala	Gln	Val
					35				40				45	
Gly	Ala	Pro	Ala	Ser	Ser	Cys	Ala	Pro	Arg	Gln	Asp	Leu	Ser	Ala
					50				55				60	
Pro	Arg	Gly	Ala	Ala	Val	Gly	Val	Ala	Gln	Ala	Glu	Ser	Gly	Pro
					65				70				75	
Ala	Pro	Ala	Arg	Val	Val	Pro	Ser	Ala	Ala	Glu	Thr	Cys	Arg	Ala
					80				85				90	
Pro	Gln	Arg	Gly	His	Gly	Ala	Ala	Gln	Ser	Ala	Val	Thr	Val	Arg
					95				100				105	
Gln	Arg	Gly	Gly	Leu	Gly	Arg	Gly	Ser	Ala	Leu	Ser	Pro	Gln	Arg
				110				115					120	
Arg	Pro	Gly	Arg	Ser	Thr	Arg	Ile	Leu	Ile	Pro	Ser	Ala	Ala	Pro
					125				130				135	
Ala	Arg	Gly	Ser	Leu	Leu	Ala	Leu	Thr	Phe	Arg	Ser	Cys	Arg	Phe
				140				145					150	
Lys	Asn	Ser	Ile	Cys	Ser	Gly	Ser	Ser	Arg	Arg	Gly	Glu	Gly	Asp
				155				160					165	
Cys	Pro	Ala	Leu	Ala	Arg	Cys	Gly	Ala	Ala	Pro	Val	Pro	Thr	Cys
				170				175					180	
Gly	Arg	Ser	Pro	Phe	Pro	Ala	Ala	Leu	Arg	Gly	Pro	Pro	Ala	Arg
				185				190					195	
Arg	Ala	Gly	Lys	Ala	Ala	Arg	Glu	Ala	Gly	Gly	Leu	Gly	Glu	Thr
				200				205					210	
Gln	Ala	Leu	Cys	Arg	Glu	Asp	Ala	Ile	Arg	Gly	Gly	Ala	Thr	Cys
				215				220					225	
Ala	Arg	Lys	Arg	Glu	Leu	Gly	Ser	Pro	Gly	Trp	Asn	Ala	Ala	Asp
				230				235					240	
Thr	Cys	Pro	Glu	Val	Pro	Glu	Tyr	Gly	Thr	Leu	Val	His	Pro	Arg
				245				250					255	
Ala	Met	Leu	Ser	Pro	Asp	Leu	Gln	Ala	Leu	Gly	Trp	Ala	Leu	Lys
				260				265					270	
Ala	Met	Leu	Arg	Leu	Gln	Glu	Ile	Arg	Glu	Ala	Phe	Lys	Val	Phe
				275				280					285	
Asp	Arg	Asp	Gly	Asn	Gly	Phe	Ile	Ser	Lys	Gln	Glu	Leu	Gly	Thr
				290				295					300	
Ala	Met	Arg	Ser	Leu	Gly	Tyr	Met	Pro	Asn	Glu	Val	Glu	Leu	Glu
				305				310					315	
Val	Ile	Ile	Gln	Arg	Leu	Asp	Met	Asp	Gly	Asp	Gly	Gln	Val	Asp
				320				325					330	
Phe	Glu	Glu	Phe	Val	Thr	Leu	Leu	Gly	Pro	Lys	Leu	Ser	Thr	Ser
				335				340					345	
Gly	Ile	Pro	Glu	Lys	Phe	His	Gly	Thr	Asp	Phe	Asp	Thr	Val	Phe
				350				355					360	
Trp	Lys	Cys	Asp	Met	Gln	Lys	Leu	Thr	Val	Asp	Glu	Leu	Lys	Arg
				365				370					375	
Leu	Leu	Tyr	Asp	Thr	Phe	Cys	Glu	His	Leu	Ser	Met	Lys	Asp	Ile
				380				385					390	
Glu	Asn	Ile	Ile	Met	Thr	Glu	Glu	Ser	His	Leu	Gly	Thr	Ala	
				395				400					405	
Glu	Glu	Cys	Pro	Val	Asp	Val	Glu	Thr	Cys	Ser	Asn	Gln	Gln	Ile
				410				415					420	
Arg	Gln	Thr	Cys	Val	Arg	Lys	Ser	Leu	Ile	Cys	Ala	Phe	Ala	Ile
				425				430					435	
Ala	Phe	Ile	Ile	Ser	Val	Met	Leu	Ile	Ala	Ala	Asn	Gln	Val	Leu
				440				445					450	
Arg	Ser	Gly	Met	Lys										
				455										

<210> 7
 <211> 407
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5583544CD1

<400> 7

Met	Leu	Phe	Asp	Lys	Val	Lys	Ala	Phe	Ser	Val	Gln	Leu	Asp	Gly
1					5					10				15
Ala	Thr	Ala	Gly	Val	Glu	Pro	Val	Phe	Ser	Gly	Gly	Gln	Ala	Val
					20					25				30
Ala	Gly	Arg	Val	Leu	Leu	Glu	Leu	Ser	Ser	Ala	Ala	Arg	Val	Gly
					35					40				45
Ala	Leu	Arg	Leu	Arg	Ala	Arg	Gly	Arg	Ala	His	Val	His	Trp	Thr
					50					55				60
Glu	Ser	Arg	Ser	Ala	Gly	Ser	Ser	Thr	Ala	Tyr	Thr	Gln	Ser	Tyr
					65					70				75
Ser	Glu	Arg	Val	Glu	Val	Val	Ser	His	Arg	Ala	Thr	Leu	Leu	Ala
					80					85				90
Pro	Asp	Thr	Gly	Glu	Thr	Thr	Thr	Leu	Pro	Pro	Gly	Arg	His	Glu
					95					100				105
Phe	Leu	Phe	Ser	Phe	Gln	Leu	Pro	Pro	Thr	Leu	Val	Thr	Ser	Phe
					110					115				120
Glu	Gly	Lys	His	Gly	Ser	Val	Arg	Tyr	Cys	Ile	Lys	Ala	Thr	Leu
					125					130				135
His	Arg	Pro	Trp	Val	Pro	Ala	Arg	Arg	Ala	Arg	Lys	Val	Phe	Thr
					140					145				150
Val	Ile	Glu	Pro	Val	Asp	Ile	Asn	Thr	Pro	Ala	Leu	Leu	Ala	Pro
					155					160				165
Gln	Ala	Gly	Ala	Arg	Glu	Lys	Val	Ala	Arg	Ser	Trp	Tyr	Cys	Asn
					170					175				180
Arg	Gly	Leu	Val	Ser	Leu	Ser	Ala	Lys	Ile	Asp	Arg	Lys	Gly	Tyr
					185					190				195
Thr	Pro	Gly	Glu	Val	Ile	Pro	Val	Phe	Ala	Glu	Ile	Asp	Asn	Gly
					200					205				210
Ser	Thr	Arg	Pro	Val	Leu	Pro	Arg	Ala	Ala	Val	Val	Gln	Thr	Gln
					215					220				225
Thr	Phe	Met	Ala	Arg	Gly	Ala	Arg	Lys	Gln	Lys	Arg	Ala	Val	Val
					230					235				240
Ala	Ser	Leu	Ala	Gly	Glu	Pro	Val	Gly	Pro	Gly	Gln	Arg	Ala	Leu
					245					250				255
Trp	Gln	Gly	Arg	Ala	Leu	Arg	Ile	Pro	Pro	Val	Gly	Pro	Ser	Ile
					260					265				270
Leu	His	Cys	Arg	Val	Leu	His	Val	Asp	Tyr	Ala	Leu	Lys	Val	Cys
					275					280				285
Val	Asp	Ile	Pro	Gly	Thr	Ser	Lys	Leu	Leu	Leu	Glu	Leu	Pro	Leu
					290					295				300
Val	Ile	Gly	Thr	Ile	Pro	Leu	His	Pro	Phe	Gly	Ser	Arg	Ser	Ser
					305					310				315
Ser	Val	Gly	Ser	His	Ala	Ser	Phe	Leu	Leu	Asp	Trp	Arg	Leu	Gly
					320					325				330
Ala	Leu	Pro	Glu	Arg	Pro	Glu	Ala	Pro	Pro	Glu	Tyr	Ser	Glu	Val
					335					340				345
Val	Ala	Asp	Thr	Glu	Glu	Ala	Ala	Leu	Gly	Gln	Ser	Pro	Phe	Pro
					350					355				360
Leu	Pro	Gln	Asp	Pro	Asp	Met	Ser	Leu	Glu	Gly	Pro	Phe	Phe	Ala
					365					370				375
Tyr	Ile	Gln	Glu	Phe	Arg	Tyr	Arg	Pro	Pro	Pro	Leu	Tyr	Ser	Glu
					380					385				390
Glu	Asp	Pro	Asn	Pro	Leu	Leu	Gly	Asp	Met	Arg	Pro	Arg	Cys	Met
					395					400				405
Thr	Cys													

<210> 8
 <211> 240
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3459449CD1

<400> 8
 Met Ala Leu Ala Ala Leu Met Ile Ala Leu Gly Ser Leu Gly Leu
 1 5 10 15
 His Thr Trp Gln Ala Gln Ala Val Pro Thr Ile Leu Pro Leu Gly
 20 25 30
 Leu Ala Pro Asp Thr Phe Asp Asp Thr Tyr Val Gly Cys Ala Glu
 35 40 45
 Glu Met Glu Glu Lys Ala Ala Pro Leu Leu Lys Glu Glu Met Ala
 50 55 60
 His His Ala Leu Leu Arg Glu Ser Trp Glu Ala Ala Gln Glu Thr
 65 70 75
 Trp Glu Asp Lys Arg Arg Gly Leu Thr Leu Pro Pro Gly Phe Lys
 80 85 90
 Ala Gln Asn Gly Ile Ala Ile Met Val Tyr Thr Asn Ser Ser Asn
 95 100 105
 Thr Leu Tyr Trp Glu Leu Asn Gln Ala Val Arg Thr Gly Gly
 110 115 120
 Ser Arg Glu Leu Tyr Met Arg His Phe Pro Phe Lys Ala Leu His
 125 130 135
 Phe Tyr Leu Ile Arg Ala Leu Gln Leu Leu Arg Gly Ser Gly
 140 145 150
 Cys Ser Arg Gly Pro Gly Glu Val Val Phe Arg Gly Val Gly Ser
 155 160 165
 Leu Arg Phe Glu Pro Lys Arg Leu Gly Asp Ser Val Arg Leu Gly
 170 175 180
 Gln Phe Ala Ser Ser Ser Leu Asp Lys Ala Val Ala His Arg Phe
 185 190 195
 Gly Glu Lys Arg Arg Gly Cys Val Ser Ala Pro Gly Val Gln Leu
 200 205 210
 Gly Ser Gln Ser Glu Gly Ala Ser Ser Leu Pro Pro Trp Lys Thr
 215 220 225
 Leu Leu Leu Ala Pro Gly Glu Phe Gln Leu Ser Gly Val Gly Pro
 230 235 240

<210> 9
 <211> 97
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2544505CD1

<400> 9
 Met Ala Leu Leu Ala Leu Ala Ser Ala Val Pro Ser Ala Leu Leu
 1 5 10 15
 Ala Leu Ala Val Phe Arg Val Pro Ala Trp Ala Cys Leu Leu Cys
 20 25 30
 Phe Thr Thr Tyr Ser Glu Arg Leu Arg Ile Cys Gln Met Phe Val
 35 40 45
 Gly Met Arg Ser Pro Ser Leu Lys Ser Val Arg Arg Pro Ser Arg
 50 55 60
 Pro Pro Ser Arg Ala Ser Leu Thr Pro Lys Ser Val Arg Arg Pro
 65 70 75
 Ser Thr Leu His Gln Cys Pro Gly Glu Gly Ala Glu Gly Gly Gln
 80 85 90
 Glu Arg Pro Arg Gly Ser Gly

<210> 10
 <211> 1161
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1568561CD1

<400> 10
 Met Trp Ala Met Glu Ser Gly His Leu Leu Trp Ala Leu Leu Phe
 1 5 10 15
 Met Gln Ser Leu Trp Pro Gln Leu Thr Asp Gly Ala Thr Arg Val
 20 25 30
 Tyr Tyr Leu Gly Ile Arg Asp Val Gln Trp Asn Tyr Ala Pro Lys
 35 40 45
 Gly Arg Asn Val Ile Thr Asn Gln Pro Leu Asp Ser Asp Ile Val
 50 55 60
 Ala Ser Ser Phe Leu Lys Ser Asp Lys Asn Arg Ile Gly Gly Thr
 65 70 75
 Tyr Lys Lys Thr Ile Tyr Lys Glu Tyr Lys Asp Asp Ser Tyr Thr
 80 85 90
 Asp Glu Val Ala Gln Pro Ala Trp Leu Gly Phe Leu Gly Pro Val
 95 100 105
 Leu Gln Ala Glu Val Gly Asp Val Ile Leu Ile His Leu Lys Asn
 110 115 120
 Phe Ala Thr Arg Pro Tyr Thr Ile His Pro His Gly Val Phe Tyr
 125 130 135
 Glu Lys Asp Ser Glu Gly Ser Leu Tyr Pro Asp Gly Ser Ser Gly
 140 145 150
 Pro Leu Lys Ala Asp Asp Ser Val Pro Pro Gly Gly Ser His Ile
 155 160 165
 Tyr Asn Trp Thr Ile Pro Glu Gly His Ala Pro Thr Asp Ala Asp
 170 175 180
 Pro Ala Cys Leu Thr Trp Ile Tyr His Ser His Val Asp Ala Pro
 185 190 195
 Arg Asp Ile Ala Thr Gly Leu Ile Gly Pro Leu Ile Thr Cys Lys
 200 205 210
 Arg Gly Ala Leu Asp Gly Asn Ser Pro Pro Gln Arg Gln Asp Val
 215 220 225
 Asp His Asp Phe Phe Leu Leu Phe Ser Val Val Asp Glu Asn Leu
 230 235 240
 Ser Trp His Leu Asn Glu Asn Ile Ala Thr Tyr Cys Ser Asp Pro
 245 250 255
 Ala Ser Val Asp Lys Glu Asp Glu Thr Phe Gln Glu Ser Asn Arg
 260 265 270
 Met His Ala Ile Asn Gly Phe Val Phe Gly Asn Leu Pro Glu Leu
 275 280 285
 Asn Met Cys Ala Gln Lys Arg Val Ala Trp His Leu Phe Gly Met
 290 295 300
 Gly Asn Glu Ile Asp Val His Thr Ala Phe Phe His Gly Gln Met
 305 310 315
 Leu Thr Thr Arg Gly His His Thr Asp Val Ala Asn Ile Phe Pro
 320 325 330
 Ala Thr Phe Val Thr Ala Glu Met Val Pro Trp Glu Pro Gly Thr
 335 340 345
 Trp Leu Ile Ser Cys Gln Val Asn Ser His Phe Arg Asp Gly Met
 350 355 360
 Gln Ala Leu Tyr Lys Val Lys Ser Cys Ser Met Ala Pro Pro Val
 365 370 375
 Asp Leu Leu Thr Gly Lys Val Arg Gln Tyr Phe Ile Glu Ala His
 380 385 390
 Glu Ile Gln Trp Asp Tyr Gly Pro Met Gly His Asp Gly Ser Thr
 395 400 405
 Gly Lys Asn Leu Arg Glu Pro Gly Ser Ile Ser Asp Lys Phe Phe

410	415	420
Gln Lys Ser Ser Ser Arg Ile Gly Gly	Thr Tyr Trp Lys Val Arg	
425	430	435
Tyr Glu Ala Phe Gln Asp Glu Thr Phe	Gln Glu Lys Met His Leu	
440	445	450
Glu Glu Asp Arg His Leu Gly Ile Leu	Gly Pro Val Ile Arg Ala	
455	460	465
Glu Val Gly Asp Thr Ile Gln Val Val	Phe Tyr Asn Arg Ala Ser	
470	475	480
Gln Pro Phe Ser Met Gln Pro His Gly	Val Phe Tyr Glu Lys Asp	
485	490	495
Tyr Glu Gly Thr Val Tyr Asn Asp Gly	Ser Ser Tyr Pro Gly Leu	
500	505	510
Val Ala Lys Pro Phe Glu Lys Val Thr	Tyr Arg Trp Thr Val Pro	
515	520	525
Pro His Ala Gly Pro Thr Ala Gln Asp	Pro Ala Cys Leu Thr Trp	
530	535	540
Met Tyr Phe Ser Ala Ala Asp Pro Ile	Arg Asp Thr Asn Ser Gly	
545	550	555
Leu Val Gly Pro Leu Leu Val Cys Arg	Ala Gly Ala Leu Gly Ala	
560	565	570
Asp Gly Lys Gln Lys Gly Val Asp Lys	Glu Phe Phe Leu Leu Phe	
575	580	585
Thr Val Leu Asp Glu Asn Lys Ser Trp	Tyr Ser Asn Ala Asn Gln	
590	595	600
Ala Ala Ala Met Leu Asp Phe Arg Leu	Leu Ser Glu Asp Ile Glu	
605	610	615
Gly Phe Gln Asp Ser Asn Arg Met His	Ala Ile Asn Gly Phe Leu	
620	625	630
Phe Ser Asn Leu Pro Arg Leu Asp Met	Cys Lys Gly Asp Thr Val	
635	640	645
Ala Trp His Leu Leu Gly Leu Gly Thr	Glu Thr Asp Val His Gly	
650	655	660
Val Met Phe Gln Gly Asn Thr Val Gln	Leu Gln Gly Met Arg Lys	
665	670	675
Gly Ala Ala Met Leu Phe Pro His Thr	Phe Val Met Ala Ile Met	
680	685	690
Gln Pro Asp Asn Leu Gly Thr Phe Glu	Ile Tyr Cys Gln Ala Gly	
695	700	705
Ser His Arg Glu Ala Gly Met Arg Ala	Ile Tyr Asn Val Ser Gln	
710	715	720
Cys Pro Gly His Gln Ala Thr Pro Arg	Gln Arg Tyr Gln Ala Ala	
725	730	735
Arg Ile Tyr Tyr Ile Met Ala Glu Glu	Val Glu Trp Asp Tyr Cys	
740	745	750
Pro Asp Arg Ser Trp Glu Arg Glu Trp	His Asn Gln Ser Glu Lys	
755	760	765
Asp Ser Tyr Gly Tyr Ile Phe Leu Ser	Asn Lys Asp Gly Leu Leu	
770	775	780
Gly Ser Arg Tyr Lys Lys Ala Val Phe	Arg Glu Tyr Thr Asp Gly	
785	790	795
Thr Phe Arg Ile Pro Arg Pro Arg Thr	Gly Pro Glu Glu His Leu	
800	805	810
Gly Ile Leu Gly Pro Leu Ile Lys Gly	Glu Val Gly Asp Ile Leu	
815	820	825
Thr Val Val Phe Lys Asn Asn Ala Ser	Arg Pro Tyr Ser Val His	
830	835	840
Ala His Gly Val Leu Glu Ser Thr Thr	Val Trp Pro Leu Ala Ala	
845	850	855
Glu Pro Gly Glu Val Val Thr Tyr Gln	Trp Asn Ile Pro Glu Arg	
860	865	870
Ser Gly Pro Gly Pro Asn Asp Ser Ala	Cys Val Ser Trp Ile Tyr	
875	880	885
Tyr Ser Ala Val Asp Pro Ile Lys Asp	Met Tyr Ser Gly Leu Val	
890	895	900
Gly Pro Leu Ala Ile Cys Gln Lys Gly	Ile Leu Glu Pro His Gly	
905	910	915

Gly Arg Ser Asp Met Asp Arg Glu Phe Ala Leu Leu Phe Leu Ile
 920 925 930
 Phe Asp Glu Asn Lys Ser Trp Tyr Leu Glu Glu Asn Val Ala Thr
 935 940 945
 His Gly Ser Gln Asp Pro Gly Ser Ile Asn Leu Gln Asp Glu Thr
 950 955 960
 Phe Leu Glu Ser Asn Lys Met His Ala Ile Asn Gly Lys Leu Tyr
 965 970 975
 Ala Asn Leu Arg Gly Leu Thr Met Tyr Gln Gly Glu Arg Val Ala
 980 985 990
 Trp Tyr Met Leu Ala Met Gly Gln Asp Val Asp Leu His Thr Ile
 995 1000 1005
 His Phe His Ala Glu Ser Phe Leu Tyr Arg Asn Gly Glu Asn Tyr
 1010 1015 1020
 Arg Ala Asp Val Val Asp Leu Phe Pro Gly Thr Phe Glu Val Val
 1025 1030 1035
 Glu Met Val Ala Ser Asn Pro Gly Thr Trp Leu Met His Cys His
 1040 1045 1050
 Val Thr Asp His Val His Ala Gly Met Glu Thr Leu Phe Thr Val
 1055 1060 1065
 Phe Ser Arg Thr Glu His Leu Ser Pro Leu Thr Val Ile Thr Lys
 1070 1075 1080
 Glu Thr Glu Lys Ala Val Pro Pro Arg Asp Ile Glu Glu Gly Asn
 1085 1090 1095
 Val Lys Met Leu Gly Met Gln Ile Pro Ile Lys Asn Val Glu Met
 1100 1105 1110
 Leu Ala Ser Val Leu Val Ala Ile Ser Val Thr Leu Leu Leu Val
 1115 1120 1125
 Val Leu Ala Leu Gly Gly Val Val Trp Tyr Gln His Arg Gln Arg
 1130 1135 1140
 Lys Leu Arg Arg Asn Arg Arg Ser Ile Leu Asp Asp Ser Phe Lys
 1145 1150 1155
 Leu Leu Ser Phe Lys Gln
 1160

<210> 11
 <211> 1806
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2508226C81

<400> 11
 gtccctgagt aggtgaggag gtgggttagga gcttgcattat agaaaaagtgg aatcgagtag 60
 tccttgctgg tggagccgct gcccgcagg aactcaggc cggctcctgt tccttcaaga 120
 gtgctggagg ccaaacttga aatacagaat taatgttccct cgtcggggcaa aagataagga 180
 tccgatctcc cccggccccc tgcgcaggcag gagcgcaccaa cccgcacccg gttaaaaact 240
 cccaggact ctgcgtgtc gccacctttt gttctctccc ccgttccac tcggggcttc 300
 cctcaggggcc gggaggcaca gcggtccctg cttgcgtggg ggctggatgt acgcattccgc 360
 aggttcccgc ggacttgggg ggcgcgtc agccccggcg cccgcagaag acttgtttt 420
 gcctcttgcgac gcttcacccccc ggaggccgc gaggccctac caccatgtac actgggtgt 480
 tcagcatgcg cttgtggacc ccagtggcg tcctgaccc tcgtggctac tgcctgcacc 540
 agcgggggt gcccctggcc gagctgcagg aggccgatgg ccagtgtccg gtcgaccgca 600
 gcctgtgaa gtggaaaatg gtgcaggctg tggttcgaca cggggctcgg agtcccttca 660
 agccgcgtcc gctggaggag caggtagat ggaacccccc gctttagag gtcccacccc 720
 aaactcagg tggatcacaca gtcaccaatc tagctgggg tccgaaacca tattctcctt 780
 acgactctca ataccatgag accaccctga agggggcat gtttgcgtgg cagctgacca 840
 aggtggccat gcagcaaatg tttgccttgg gagagagact gaggaagaac tatgtgaaag 900
 acattccctt tctttcacca accttcaccc cacaggaggt ctttattcgt tccactaaca 960
 tttttcgaa tctggagttc acccggttgc tgctggctgg gctttccag tgtcagaaag 1020
 aaggacccat catcatccac actgtatggc cagattcaga agtcttgcgtat cccaaactacc 1080
 aaagctgtc gaggctgagg cagagaacca gaggccggag gcagactgc tctttacagc 1140
 caggaatctc agaggattt gaaaagggtga aggacaggat gggcattgac agtagtgata 1200
 aagtggactt cttcatctc ctggacaacg tggctgcccga gcaggcacac aacctcccaa 1260
 gctgccccat gctgaagaga tttgcacgga tgatcgaaca gagagctgtg gacacatctt 1320

tgtacatact	gcccaaggaa	gacagggaaa	gtcttcagat	ggcagtaggc	ccattcctcc	1380
acatccctaga	gagcaacctg	ctgaaagcca	tggactctgc	caactgcccc	gacaagatca	1440
gaaagctgtt	tctctatgcg	gctcatgtat	tgaccttcat	accgcttta	atgaccctgg	1500
ggattttta	ccacaaatgg	ccaccgtttt	ctgttgcac	gaccatggaa	ctttaccacgc	1560
accttggaaatc	taaggagtgg	tttgcgcgc	tctattacca	cgggaaaggag	cagggtgcgcg	1620
gagggttgcgc	tgatgggcct	tgcggctgg	acatgtttt	gaatggccat	tcaagtttata	1680
ccttaagccc	agaaaaataac	cacgcactct	gcttcacaaac	tcaggtgtat	gaagtggaa	1740
atgaagagta	actgatttat	aaaagcagga	tgtgttgcatt	ttaaaaataaa	gtgcctttat	1800
acaatg						1806

<210> 12
<211> 1271

<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 2880964CB1

<400> 12

<210> 13
<211> 1402

<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 3122091CB1

<400> 13

acggggagcg	ggagccggag	ccggagcggg	acccggagcg	agagcgcgcg	ggcgcggccc	60
acagtgcctg	atttgagatg	gggtcccgagg	tctcggtgg	atcgggagct	ctgcacgtgg	120
tgtatgtggg	tgggggcttt	ggccggatcg	cagcagccag	ccagctgcag	gccctgaacag	180
tcccccatt	gtctgggac	atgaaggact	cttccacca	caatgtggt	gtctccggag	240
ccttcgttga	gacagggttc	gccaanaaaaga	cattcattc	ttactcggt	atttcaagg	300
acaacttccg	cgacggggcta	gttagtgggaa	tagacctgaa	gaaccatag	gtgctgtgc	360
agggtggcga	ggccctgccc	ttctctcatc	ttatcctggc	cacgggcagc	actggggccct	420
tccccggca	gttaatgag	gttccagcc	agcaggccgc	tatccaggcc	tatgaggaca	480
tggtgaggca	ggtccagcgc	tcacggttca	tcgtgggtt	gggaggaggc	tcggctggag	540
tggagatggc	agcagagatt	aaaacagaat	atccctgagaa	agaggtaact	ctcattcact	600
cccaagtggc	cctggctgac	aggaggctcc	tgcctccctgt	ccggcaggaa	gtgaaggaga	660
tcctcctccg	gaaggggctg	cagctgtgc	tgatgtggc	ggtagagcc	ctggaggagc	720
tgcccttcata	tgatgtatcg	gagtagatca	aagtgcagac	ggacaaaaggc	acagaggatgg	780
ccaccaacct	ggtgattctc	tgacccggca	tcaagatcaa	cagctccggcc	taccgc当地	840

cattttagatc cagactatcc agcagtgtt ctctgatgtt gaacgagcac ctccaggtgg 900
 agggccacag caacgtctac gccattgtt actgtgccg cgtgaggacg cccaaatgg 960
 cctatcttgc cggccctccac gccaacatcg ccgtggccaa catcgtaac tctgtaaac 1020
 agcggccctc ccaggctac aagccgggtt cactgacgtt cctctgtcc atggggagaa 1080
 atgacggtgtt gggccaaatc agtggcttct atgtggccg gctcatgggtt cggctgacca 1140
 agagccggta cctgttcgtc tctacgactt gggaaaccat gaggcagtctt ccaccttgat 1200
 ggagaggcca ggcgggagaa ctaccgcgc aggtggcggt acggactgtt tggcgatgg 1260
 caccggctc gcaatgtcta gaactaatgc tattttctg gaataagatg ccaatgtat 1320
 ggtggctaga aatgcaactt gtataaaaca aaaatgggag agagagaggtt attaaacaaa 1380
 taccggccctt aaaaaaaaaaa aa 1402

<210> 14
 <211> 1156
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1633313CB1

<400> 14
 gggaggcctg acctgaggcc ggcacccgga gctggcgaaa gcagacccag aggaccccg 60
 cccgttcttgc ctgtcagcag cctggccggag cttcccacag tggctgtctg gagggtgtgg 120
 cagccaaatgc tccagctgtt cagataactga tgtccctac ctgctgtttgg ccgtcaatgc 180
 agaacccaggcg cgttgcag aacgacaggc cgtgagagag acgtggggca gtccagctcc 240
 agggatccgg cgtcttcc tctgtgggtc tccggtaggtt gggccggggc ctgacccatgg 300
 ctcacttagtgc gcatgggaga gccgtcgatc cagcgacctg ctgctctggg accttcctcg 360
 cgtcccatcc aaccagacgc tcaaaagaccc tctgtgtctg gcctggctgg gcccactg 420
 ccccaaccgtt agttttgttct tgcagatcga ggacatgtcc ttgttacaca cccctggcc 480
 gctggctcac ctggggggcc tggccacatgc ctggccggaa agccatctacc tgggtgagg 540
 ctttaccccg gcatgtcc tccggaaagcc aggaggaccc ttcataatgtgc ccgagtcctt 600
 cttcgaaatgtt ggttaccccg cctatgtcaag cgggggtggc tacgttattt ccggccctt 660
 ggcacccctgg ctgtgtgggg cggcagcccg tttggcaccctt tttcccttttgggaggctt 720
 cactggccctt tgcattcccgag ccctggccctt ggtggccctt gcccacccag gtttccctcac 780
 agcctggccca gcaagaccgc ctggggacca ctgtgttttgc cggcaacccgtc tgctggtag 840
 gcccctggc ccccaggccca gcatccgtt ctggaaaacaa ctgcaagacc caaggctcca 900
 gtgtctacttcc tcatggggaa gggccggaggtt gctgacccgtt ccggccctt ggcctgggg 960
 tctggggccg gcccctggctt cagggcccttcc ttccagggttct tgatggggagg gaggaggcc 1020
 cagaagctgg acaacttaag ccactcccttgc ttccatggcc gccaggggcc tggccaggaa 1080
 agatgggtt gttggactgtt ttgttactt gaaaaacatgtt cactccctcac 1140
 tctgaaaaaa aaaaaaa 1156

<210> 15
 <211> 3908
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2323763CB1

<400> 15
 gtccaggggag gctgaggcga gaggtatgtt tccgggtggg gagcccgac taccttctt 60
 ctcttcctcc tccctctccg ggtggggaa gcgaaagggtt ggggtccccc agccatgg 120
 ccaggaggag gcccggggccg ccggaggccg gggcccccgtt atgtggccctt gagcccccgtt 180
 tactgtttctt gctgtctgg agggccatgg agaagaggctt gggagtcaag ccaatccctg 240
 cttccctggat ttatcagga tattattggc agacatctgc gaagtgggtt agaaggctgt 300
 acctgtttt tactgtttt tgcttccatgc ttctgtgggtt gtcaacatgc gccagtggata 360
 gcagggtccca gcaaggggaaag acacaatgtt ggtggccctt gagatgttgggggg 420
 acctctggctt tcttgaagga accccccttc tccatgtatgc ttgggtgtcc tgctggccagg 480
 actctgtcttccatgtctt ttgggtgttccatgtatgc tttggatgttccatgtatgc gactgcacca 540
 gggcccccacatgc tttggatgttccatgtatgc tttggatgttccatgtatgc gactgcacca 600
 aaaaatttcca aactgcacatgtt gatggggctt ttcataatgttca agatgtatgttca ccacatcttc 660
 tggggcttggggatgtt gcatcttggaa ggcaggccccc accccagatgttccatgtatgc gactgcacca 720
 ctgtgttccatgtatgc tttggatgttccatgtatgc tttggatgttccatgtatgc gactgcacca 780
 ccaggatgttccatgtatgc tttggatgttccatgtatgc tttggatgttccatgtatgc gactgcacca 840
 taggtgggttccatgtatgc gactaccatgttccatgtatgc gactaccatgttccatgtatgc gactaccatgttccatgtatgc 900

ccctaaccac agacctgact gcagagctgt ctgggggcc aaagaatgta tcagtgcac 960
 ctgaaatc agagggtctt gctactacgc ccageactca acaagtaaaa agttctgaga 1020
 aaaccagat tgctgtcccc cagccagttt ctccctccta cagttatgct acccctaccc 1080
 cccagggctc tttccagagc acctcagcac catacccagt tataaaggaa ctgggttat 1140
 ctgctggaga gagggtccag ataaccctgc ctaagaatgta agttcaatta aatgcatacg 1200
 ttctccaaga accacctaaa ggagaaacct acacccatcg ctggcagctg attactcatc 1260
 cttagagacta cagtggagaa atggaaggg aacatccca gatccctcaa ctatcgaagc 1320
 tcactccagg cctgtatgaa ttcaaaatgta tttagaggg tcaaaatgcc catgggaag 1380
 gctatgtgaa cgtgacagtc aagccagagc cccgtaaagaa tcggcccccc attgctattg 1440
 tgcacccca gttccaggag atcttttgc caaccacttc tacagtcatt gatggcagtc 1500
 aaagcaactga tggatgataaa atcggtcaat accatggga agaacttaag gggcctctaa 1560
 gagaagagaa gatttctgaa gatacagcca tattaaaact aagttaaactc gtccctggg 1620
 actacactt cagtttgcgtatgatgact ctgatggcgt tccaaactct actactgcaa 1680
 acctgacagt gaacaaatggtt gttgattacc cccctgtggc caacgcaggc cccaaacaag 1740
 tggatcaccct gccccaaaac tccatcaccc tctttggaa ccagagcact gatgatcatg 1800
 gcatcaccag ctatgagtgg tcactcagcc caagcagcaa agggaaatgt gtggagatgc 1860
 aggggttttag aacaccaacc ttacagctt ctgcgtatgca agaaggagac tacacttacc 1920
 agctcacaatg gactgacaca ataggacagc agggcactgc tcaagtgtact gttattgtgc 1980
 aacctgtaaaa caataagccct cctcaggcag atcgaggccc agataaaagag ctgacccttc 2040
 ctgtggatag cacaacccctg gatggcagca agagctcaga tgatcagaaa attatctcat 2100
 atctctggaa aaaaacacag ggacctgtat ggggtcagct cgagaatgt aacagcagtg 2160
 ttgtactgt gactgggctg caagtgggaa cctatgtttt caccttgcgt gtcaaaagatg 2220
 agaggaacct gcaaaagccag agctctgtt atgtcattgtt caaagaagaa ataaacaaac 2280
 cacctatacg caagataact gggatgtt tgattaccc accacgcaggc acagcagagc 2340
 tggatggctc taatgttca gatgacaaggaa gaatagttag ctacctctgg actcggatgt 2400
 aggggagccc agcagcagggtt gagggtttaa atactctga ccattacccct atcccttttc 2460
 tttcaaaacct ggttggggaa acctacactt ttcacccgtt agtgaccgt gcaaagggtg 2520
 agagtgacac agaccggacc actgtggagg tggaaacctga tcccaggaaa aacaacctgg 2580
 tggagatcat ctggatatac aacgtcagtc agctaactgaa ggggtgaaag gggatgttca 2640
 tccggccagat tgggttctc ctgggggtgc tggattccga catcattgtt caaaagattc 2700
 agcgttacac ggagcagagc accaaaatgg tatttttttgc tcaaaacagcg cctcccccacc 2760
 agatctcaa agggcatgag gtggcagcga tgctcaagag tgagctgcgg aagcaaaagg 2820
 cagactttt gatattcaga gccttggaaag tcaacactgt cacatgtcag ctgaactgtt 2880
 ccgaccatgg ccactgtgac tgcttccaca aacgtctgtat ctgtgaccct ttttggatgg 2940
 agaatttcat caagggtcagc ctgagggatg gagacagcaa ctgtgatgtt acgtgttat 3000
 atgttacat tgcatttttgc tgcattgtt ttgccttggg aatcccttgc tggactgtga 3060
 tctgttgggtt taaggaggcaaa agggaaaaccc ccaagggaaa aagcaagttac aagatcttgg 3120
 atgccacggg tcagggaaagc ctggagctga agccaaacctc ccggagcaggc atcaaacaga 3180
 aaggcccttt gcttaatgtac agcctgtatc actccggatc agagctggac agcgatgtatg 3240
 ccatcttac atggccagac cgagagaagg gcaaaactctt gcatgttgcg aatggctctg 3300
 tacccaaacgg gcaagccctt ctgaaggcga ggagcccgcg ggaggagatc ctgtagccac 3360
 ctggctgttc tccctaggcc agggcccgac acactggccg gccagttccgc taccctcccg 3420
 agtctggggg cagctgtgtt cccagcatctt gtcgttgcatt tgccttgcg agtccaaacc 3480
 agaaccctgg ggacttgaat ccagagactt cctccaggaa cccctcaacg aagctgtgaa 3540
 tgaagagggtt tcccttttaa acctgtctgg tggggccccc gatattccca cctcaggggcc 3600
 tccctttttt tgcaaaactcc tccctccccc cgaggccaga cccagccagc tgcataagctc 3660
 tgcagctccc cagttggcagc tgcattgtt cccaggtgc tgcagggttgc ggcctgtgt 3720
 gtcgtcccgca cacttggatc caaaaccaaa cactgtggcc atgggttttgc cctctctggg 3780
 gtagagatcg ccctctcgct gggcaaaagag gaagttggcac ccctccctc accacagatg 3840
 ctgagatgtt agcatagaaa tgatggcccg cctcggccctc ccaaagtgtt gggattacag 3900
 gcgtgagc 3908

<210> 16

<211> 1368

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3759234CB1

<400> 16

atgtaccggg gcatactacac cgtgccccaaac ctgtctgtcg agcagcggccc ggtggacatc 60
 ccggaggacg agctgggggg tgagtgtccg ccgggatccc cggccggccg gcccttcata 120
 ctgtgcggcc aggtggggcgc cccagctacg agctgtcccc cgcggcaaga cctgtccgca 180
 ccccgccggc cccgggtggg ggtcgctacg gggaggagcg gcccagcccc tgcccgcggt 240
 gtcggccacgct ctcggaaac ttggccggcc cccgagcggg gtcacggggc 300

ggggtagacgg	tgcggcaacg	cgccggactg	gggcgggggt	ccgcgtgag	cccccagcgc	360
cggccggcc	ggagcacccg	catccgtatc	ccctccgccc	cgccgcggc	cggtctctgt	420
ctcgcatgt	cattccgtcc	gtgtcggtt	aaaaattcaa	tctgtctggg	cagcagcaga	480
aggggagagg	gcaactgcgg	tgttctggc	cgctgcgggg	ccgcgggggt	ccccacatgc	540
ggccgtageg	ccttccctgc	agccccgggg	ggaccccccag	ccccggcgcg	ccccggaggcg	600
gccccggagg	cgggggtct	ggggagggacc	cagggccctct	gccccggagga	cgccattcgc	660
ggaggagcca	catgtcccg	gaagagggag	ctgggcagcc	cgggatggaa	tgctgcagac	720
acctgeccag	aggccccaga	atacggcacc	cttgtgcacc	ctcgggccat	gctctcacca	780
gatctgcagg	ctctgggttg	ggctctcaag	gccccatgtca	ggtgtcggaga	gatccccggag	840
gccttcaagg	tgtttggccg	tgacggcaat	gggtttcatct	ccaaaggaggg	gttggggccgg	900
gccccatgcgt	cactgggtt	catggccaaac	gagggtggagc	tggaggatcat	atcccagcgg	960
ctggacatgg	atgggtatgg	tcaatggac	tttgaggaggt	ttgtgaccct	tctggggacc	1020
aaactctcca	cctcagggat	cccagagaag	ttccatggca	ccgactttga	tactgtcttc	1080
ttggaaatgtcg	acatgcgaaa	gtgtacgggt	gtatggatgt	agcggctgtt	ctacgacacc	1140
ttctgcgagc	acctgtccat	gaaggacata	gagaacatca	tcatgcggg	ggaggagagc	1200
cacctggcca	caggccggaga	gtgtccccgt	gtatggaga	cctgtccaa	ccagcagatc	1260
cgcccgacat	ggtgtcgca	gatgttcatc	tgccgcctcg	ccatcgccct	catcatcgt	1320
gtcatgttca	ttcccgccca	ccaggtgtcg	cgcaatgtggc	tgaatgt		1368

<210> 17
<211> 2827
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 5583544CB1

<i>17</i>	<i>aagccgagag</i>	<i>gctgaaggct</i>	<i>ctgcttaggtg</i>	<i>acctcaggcc</i>	<i>tggggccctgc</i>	<i>ctctgcaccc</i>	<i>60</i>
	<i>cttttttcc</i>	<i>tcttttacccc</i>	<i>cggggctccg</i>	<i>gagagacccag</i>	<i>cttctggcc</i>	<i>ctaagggctt</i>	<i>120</i>
	<i>cagtttctcc</i>	<i>ccgggggtccg</i>	<i>cgagggaaagg</i>	<i>gggcagccgg</i>	<i>cggtgagcag</i>	<i>gggcgttttt</i>	<i>180</i>
	<i>tattttgcct</i>	<i>cacggcttggg</i>	<i>cagggcttcg</i>	<i>cccccttgc</i>	<i>cgctattttgt</i>	<i>ggagtccctt</i>	<i>240</i>
	<i>ccccggggccgg</i>	<i>ggaagagacg</i>	<i>gtgaccccaa</i>	<i>cccccccccgg</i>	<i>cctggcgat</i>	<i>aaaagccgcg</i>	<i>300</i>
	<i>ccgacgcacg</i>	<i>gcgccccggac</i>	<i>gtactgtcc</i>	<i>ggttggtag</i>	<i>cgcgcctgcg</i>	<i>cgttgacggc</i>	<i>360</i>
	<i>gatttttgt</i>	<i>tctgaggctg</i>	<i>cagcgtcgcc</i>	<i>atcttgcgt</i>	<i>gcccgttgcg</i>	<i>gatgtccggg</i>	<i>420</i>
	<i>ccaggttccg</i>	<i>cctgtctgtt</i>	<i>gttgcaccc</i>	<i>cggacgcgat</i>	<i>gtcttcgcac</i>	<i>aaggtaaaag</i>	<i>480</i>
	<i>cgttctcggt</i>	<i>gcagttggac</i>	<i>ggggcgcacg</i>	<i>cgggcgtcga</i>	<i>ccccgttgc</i>	<i>agccggccgc</i>	<i>540</i>
	<i>aggccgtggc</i>	<i>ggggccgggt</i>	<i>ctgctggagc</i>	<i>tgtcaagcgc</i>	<i>cgcgcgtgt</i>	<i>ggtccccctg</i>	<i>600</i>
	<i>ggctgcgcgc</i>	<i>gccccggccgc</i>	<i>gcccacgtgc</i>	<i>actggacccga</i>	<i>gtcgccgc</i>	<i>gccccgtctg</i>	<i>660</i>
	<i>gcacggctt</i>	<i>cacgcagacg</i>	<i>tacagtgaac</i>	<i>gcgtggaggt</i>	<i>cgtgagccac</i>	<i>cgccgcacgc</i>	<i>720</i>
	<i>tcttggcgcc</i>	<i>agataccggg</i>	<i>gagaccacga</i>	<i>cgctccctcc</i>	<i>tggggccat</i>	<i>gatgttctgt</i>	<i>780</i>
	<i>tcagtttcca</i>	<i>gtgtcccccc</i>	<i>acccttggta</i>	<i>catttttgc</i>	<i>ggggcaaacac</i>	<i>gttagtgc</i>	<i>840</i>
	<i>gctactgtat</i>	<i>caaggccacc</i>	<i>ctgcacccgc</i>	<i>cctgggtttcc</i>	<i>agcacgcgg</i>	<i>gcaaggaaagg</i>	<i>900</i>
	<i>tgttacttgt</i>	<i>catcgagcc</i>	<i>gtggacatca</i>	<i>acacccgcage</i>	<i>cctgtggca</i>	<i>cctcaagcgg</i>	<i>960</i>
	<i>gggtctggga</i>	<i>aaagggttgc</i>	<i>cgatcttggt</i>	<i>actgttaacgg</i>	<i>tggctctagtc</i>	<i>tccctttcgg</i>	<i>1020</i>
	<i>ccaaatgtcg</i>	<i>ccgcaaggcc</i>	<i>tacaccccaag</i>	<i>gagaggctcat</i>	<i>ccctgttctt</i>	<i>gccccgatgc</i>	<i>1080</i>
	<i>acaacggctt</i>	<i>cacacgttcc</i>	<i>gtgtgtcc</i>	<i>gggcagecg</i>	<i>gtgcagacata</i>	<i>cagacgttca</i>	<i>1140</i>
	<i>tggcccgagg</i>	<i>cggccggaaaag</i>	<i>caaaaacggg</i>	<i>cagtgggtgc</i>	<i>cgcctcgcg</i>	<i>ggccgagccgg</i>	<i>1200</i>
	<i>tggggcccccgg</i>	<i>gcagccggcg</i>	<i>ctgtggcagg</i>	<i>gccccggact</i>	<i>ggggatcccc</i>	<i>ccagtgggtc</i>	<i>1260</i>
	<i>cttccatctt</i>	<i>gcactgtccgc</i>	<i>gttctacacg</i>	<i>tggactacgc</i>	<i>actcaaggtc</i>	<i>tgtgtggata</i>	<i>1320</i>
	<i>tcggcaggaaac</i>	<i>gtccaaagctg</i>	<i>ctgtgtggagc</i>	<i>tgccactgtt</i>	<i>gatccgcacc</i>	<i>attcccttgc</i>	<i>1380</i>
	<i>accctttttgg</i>	<i>cagccgttcc</i>	<i>tccagcgtgg</i>	<i>gcagccacgc</i>	<i>cagcttctcg</i>	<i>ctggactgtc</i>	<i>1440</i>
	<i>ggctggggggc</i>	<i>cttgcgggag</i>	<i>cgcccttgg</i>	<i>cttcccttgc</i>	<i>gtactcggag</i>	<i>gtgttagccg</i>	<i>1500</i>
	<i>acactgagga</i>	<i>ggcagcccttgc</i>	<i>gggcagaccc</i>	<i>cttcccccgt</i>	<i>tccggaggac</i>	<i>cccgacatgc</i>	<i>1560</i>
	<i>gccttgaagg</i>	<i>cccggttcc</i>	<i>gcttacatcc</i>	<i>aagagtccg</i>	<i>ctacegc</i>	<i>ccacccctgt</i>	<i>1620</i>
	<i>actctgagga</i>	<i>ggatccaaac</i>	<i>ccacttgc</i>	<i>gggacatgag</i>	<i>ggccgcgtgc</i>	<i>atgacttgc</i>	<i>1680</i>
	<i>gaacggcaca</i>	<i>gggacccctc</i>	<i>gaggaacaag</i>	<i>gttgcacacc</i>	<i>agcttgc</i>	<i>caccatgtact</i>	<i>1740</i>
	<i>gtggggggactg</i>	<i>gctggaccaa</i>	<i>gggctgcact</i>	<i>ccccgcgtgc</i>	<i>atcaaagtgc</i>	<i>ggaaaccaag</i>	<i>1800</i>
	<i>tctcagatgt</i>	<i>aggccggggc</i>	<i>cttccggata</i>	<i>tcacatggga</i>	<i>caagggaaa</i>	<i>ggccggctgg</i>	<i>1860</i>
	<i>aatctgactt</i>	<i>acctggaccc</i>	<i>ctgtcttgc</i>	<i>gaggcatgt</i>	<i>atgcccagt</i>	<i>cagttatccg</i>	<i>1920</i>
	<i>gagactgttt</i>	<i>aataacctgt</i>	<i>cttcccagcc</i>	<i>aattgggtgt</i>	<i>gcttgcata</i>	<i>ccttaggagcc</i>	<i>1980</i>
	<i>ttcagttctgg</i>	<i>gagaaacaga</i>	<i>gccagacata</i>	<i>gacagtcc</i>	<i>gcatcacaga</i>	<i>accagaagaa</i>	<i>2040</i>
	<i>gagacctgca</i>	<i>actgttgc</i>	<i>tccagacagg</i>	<i>aaggcagagaa</i>	<i>ggcgtcccttgc</i>	<i>ccggaaaggcc</i>	<i>2100</i>
	<i>attttagctg</i>	<i>aggcttgg</i>	<i>gtacgtatag</i>	<i>gagctcgc</i>	<i>ggcagacgaa</i>	<i>tggagaacttgc</i>	<i>2160</i>
	<i>aggctcagaga</i>	<i>aggctcagac</i>	<i>tgatgtacgt</i>	<i>ttggaaatcc</i>	<i>ccccgtttat</i>	<i>tgttgcact</i>	<i>2220</i>
	<i>gggggttcaga</i>	<i>ggggcaggtgc</i>	<i>ctcagatgttgc</i>	<i>aggccacaca</i>	<i>gtgaggcttc</i>	<i>gtgggtgaaa</i>	<i>2280</i>

ggacccagga	acgaggcgtt	caggaaagca	ggttgtcaga	gctatgtgga	gtctgtgggt	2340
ggcaggggca	gccgtccag	ctttgaaga	ctttgaaaagc	cagagattcc	ttggcgcaggc	2400
tggacttc	tggagactcc	tccaagtacc	caggggcatc	agagatgcct	gggtgttaca	2460
tggccaggg	aaccagggtt	cagggttaga	caggcaagac	cagataccca	atgtcaaaag	2520
tggaaaacat	gggctccctg	ttaaaacgtat	agaattccaa	gacagtgaca	gcattacgtc	2580
acccctgggg	acagagggtca	gcctaagggt	acacacgggg	actactgtgc	ttccggaggc	2640
tccctgtgtc	ctggaggaga	agagcattag	agggggcagc	tggacaagct	cccaactgca	2700
gagtcccccagc	cctggctggg	gcagggcccc	ggcctgggac	tcagcatttc	tgatatgcct	2760
taagaattca	ttctgttttg	tacaattatt	ttttaaaagt	aaacgtgtgg	agaaaagaaaa	2820
aaaaaaaa						2827

<210> 18
<211> 1113
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 3459449CB1

```

<400> 18
gccccctgaa accaggactc cagcacctct ggtcccgccc tcacccggac ccctggccct 60
cacgtctccc ccaggatgg cgtggcgcc ttgtatgatc gcccctggca gcctcgccct 120
ccacacccctgg caggcccagg ctgttccac catcctggcc ctggccctgg ctccagacac 180
cttgcacatg acctatgtgg gtgtgcaga ggagatggag gagaaggcag cccccctgtct 240
aaaggagggaa atggcccccac atggccctgtct gcgggaaatcc tggggaggcag cccaggagac 300
ctggggaggac aagcgctcgag ggcttacctt gccccctggc ttcaaaagcccc agaatggaaat 360
agccattatg gtctacacca actcatcgaa caccttgcac tgggagttga atcaggccgt 420
gcggacggggc ggaggctccc gggagctcta catgaggcac ttcccttca aggcctgcac 480
tttctacctg atccggggcc tcacgtgtct gggggggcgt gggggctgcac gcagggggacc 540
ttggggagggtg gtgttcccgag gtgtgggcag ctttcgtttt gaaaaaaaga ggctgggggg 600
cttcgttcccg ttggggccagt ttcccttcag ctccctgtat aaggcagtgg cccacagatt 660
ttggggagaagc aggcggggctt gtgtgtctgc gcccggggtg cagctagggtt cacaatctga 720
ggggggccccc tctctggccc ctttggaaagac ttttgcgtttt gccccctggag agttccagct 780
ctcagggggtt gggccctgaa agtccaaacat ctggccactt ggagccctgg gaacgggtga 840
ccttcataatg acgaagaggc accttccagca gcttggagaa gcaagaacat gttccggac 900
ccagcccttag cagcccttc cccaaaggcc atgttggctt gggggaggccca cagcagggtct 960
gaggggaaactc tgctatgtga tggggacttc ctgggacaaag caagaaaagt actggggcgt 1020
ccacttgatt gaacgtgtt gcaatgttga gacatggagt ttatttggagg tagctacgtt 1080
attaaatgtt attgcgtgtt gggaaaaaaa aaa 1113

```

<210> 19
<211> 569
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 2544505CB1

```

<400> 19
ctaccacaca cctggggaat tgctggcctg acttctgacc cctgactcct catacccttc 60
ctccagagca tgacattga ccaccaactg aaacctgacc tctgacccca gaccactggc 120
ccttcccccgg ccctgtggtg acttcaataa ggttactagc ttctcccttg gccttgagac 180
ccacacatggc gcccgttgg ccttggccag tgccgttccg tctggcctgc tggcccttggc 240
tgtcttcagg gtgcccgcct gggccctgtct cctctgcctc acaaacctact ctgagcgcct 300
ccgcacatcgc cagatgtttg ttggggatgcg gagcccaagc ttgaagagtg tgaggaggcc 360
ttcacggccg ccttcacggg cctctctgac accgaaatca gtgaggagac catccacact 420
tcatcgtgt cctggggaaag gtgcagaggg agggcaggag agggccagag ggtcaggctg 480
aggcagacac agagagaaac agtcagcaga gaaaggctca aagaccatga gaacaacagg 540
gattaggagac agaaagacac acccagggg 569

```

<210> 20
<211> 4427
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 1568561CB1

<400> 20
ggaaaagagg gcaccaggcc cttccccctc cctcatccctc ccatecccagt aaaccctgcc 60
aaatggaaat cctgactta atttaggaga aaggccctgt aaccagatgaa ctgactgaac 120
atggctgccc gactcagggt ggggtctgca gtgcagcatt aatggccgc tgacatgaat 180
atggtagtagt tttctcttagc aaagagtaat gtgggcatg gagtcaggccc accttccttg 240
ggctctgtgt ttcatgcagt cttgtggcc tcaactgact gatggagcca ctcgagtctc 300
ctacctggcc atccggatg tgcagttggaa ctatgcctcc aagggaaagaa atgtcatcac 360
gaaccaggct ctggacaggc acatagttggc ttccagcttc taaagatctg acaagaaccg 420
gataggggaa acctacaaga accatcatc taaagaataac aaggatgact catacagaga 480
tgaatgtggc cagctgcct gttgggctt cctggggca gtgttgcagg ctgaagtgg 540
ggatgtcatt cttattcacc taaaagaaattt tgccactcgt ccctatacca tccaccctca 600
tgggtcttc tacgagaagg actctgaagg ttccctatac ccagatggct cctctggggc 660
actgaaagct gatgactctg tccccccggg gggcagccat atctacaact ggaccattcc 720
agaaggccat gcaccaggcc atgtgcacc cagctgtctc acctggatct accattctca 780
ttagatgtct cccagagaca ttgcactgt ctaatgtggg ccttcatca cctgtaaaag 840
aggagccctg gatggaaact cccctctca acggcaggat tgagaccatg atttcttcct 900
ccttcctcgt gtggtagatg agaaccctcg ctggcatctc aatggagaaca ttgcccactt 960
ctgctcagat cctgttcag tggacaaaga agatgagaca ttccaggaga gcaataggat 1020
gcatgcaatc aatggctttt tttttggaa caatggaaattt gatgtccaca ttgtccatc 1080
acgtgtggcc tggacttgt ttggcatggg ccacactgtat gtggctaaaca tttttccatc 1140
ccatggacag atgtgtact cccgtggaca ggaaccttgg acctggttaa ttagctgcctc 1200
caccttttgatc actgtgtact tgggtggctt ggcactctac aaggtcaagt cttgtccat 1260
agtgaacagt cactttcgag atggcatgca ggcccttctt gtggacctgc ttttccatc 1320
ggcccttctt gtggacctgc tcacaggca agttcgacag tacttcattt agggccatg 1380
gattcaatgg gactatggcc cgatggggca ttagatggggactt gatggggatg actgggaga 1440
gccaggcagt atctccatg agtttttccaa gaagactcc acggcaattt gggggactt 1500
ctggaaatgt cgatatgaag ctttcaaga tgacatttc acggatcc caagagaaga ttgtccat 1560
ggaagatagg ctttccatg tccctggggcc agtggatccgg gctgagggatgg tgacaccat 1620
tcaggtgttc ttctacaacc gtgcctccca gccattcage atgcagcccc atggggatc 1680
ttatgagaaa gactatgaag gactgtgtta caatgtggc tcacatttacc ctggcttgg 1740
tgccaaggcc tttgagaaaag taacataccg ctggacatgc cccctctatg cccgtccac 1800
tgcgtcaggat cctgttgc tcaacttggat gtacttcttct getcgatgg ccataagaga 1860
cacaatttc ggcctgggtt gcccgtgtct ggtgtcagg gctggtgcct ttagttggca 1920
tggcaagcag aaagggggtgg ataaagaattt ctttcttctt ttcactgtgt gggatgagaa 1980
caagagctgg tacagcaatg ccaatcaagg agctgtatg ttggatttcc gactgttttc 2040
agaggatatt gagggtttcc aagactccaa tcggatgtcat gcattaaatg ggtttctgtt 2100
ctctaaccctg cccaggctgg acatgtgcaaa ggtgtacaca gtggccctgg acctgtctgg 2160
cctggccaca gagactgtatg tgcattggat ctgttccatgg gcaacactgt tgcaatcc 2220
gggcattggg aagggtcgac ctatgtctt tccatcatacc ttttgcattgg ccatcatgca 2280
gcctgacac cttggacat ttgagatttta ttgcaggca ggcaggccatc gagaaggcagg 2340
gatgagggca atctataatg tctccctgtt ccctggccac caagccaccc ctcgcacaa 2400
ctaccaagct gcaagaatct actatatcat ggcagaaagaa tgaaatgtggg actattgtcc 2460
tgacccgggg tggggacccggg aatggccacaa ccagtctgtg aaggacatgt acggttatcat 2520
tttctgtatc aacaaggatg ggtctctggg ttccatgatc aaaaaggatg tatttgcattc 2580
atacactgtt ggtacgttca gatccctcg gccaaggact ggaccagaag aacacttggg 2640
aatcttgggt ccacttatca aaggtgaagt tggtgatatc ctgactgtgg tattcaagaa 2700
taatgccacg cgccttact ctgtgcattgc tcatggatgt cttagaatcta ctactgtctg 2760
gccaactgtct gtcggacccgtt gtggagggtgtt cacttcatcg tggaaacatcc cagagggatc 2820
tggccctggg cccaaatgt ctgttgcattt ttctgtatc tattttctc cagtgatcatc 2880
catcaaggac atgtatagtg gctctgggg gccccttggct attcgcacaa agggccatct 2940
ggagccccc gggggacggg gtcacatggaa tcgggatattt gcatgttgc ttttgatctt 3000
tgatgaaaat aagtcttggg atttggagga aatgtggca accatgggtt cccaggatcc 3060
aggcagtatt aacccatcagg atgaaacttt ctggagagc aataaaatgc atgcaatcaa 3120
tggggaaactc tatggcaacc ttaggggttctc tccatcacc atccactttt ggtggctt 3180
gtacatgtgt gccatggggcc aagatgtggaa ttcacaccatc accactttt atgcagagag 3240
cttcctcttc cggatgtggcc aagactaccg ggcagatgtg gtggatctgt tcccgaggac 3300
ttttgaggtt gtggagatgg tggccagcaaa ccctggggaca tgggtgtatgc actgcccattgt 3360
gactgaccat gtccatgtgt gcatggagac ctttttcaact gtttttctc gaacagaaca 3420
cttaaggccct ctcaccgtca tcacccaaaga gactgaaaaaa gcaatggcccccc 3480
tgaagaaggc aatgtgaaga tgcgtggccat gcatggatccc ataaagaatgtt tgagatgtct 3540
ggccctctgtt ttgggtggcc ttagtgcac ctttctgtct gttttctgg ctcttgg 3600
agtgggttgg taccacatc gacagagaaa gtcacgacgc aataggatgtt ccatcttgg 3660
tgacagcttc aagcttctgt ctttccaaaca gtaacatctg gaggctggagatcatc 3720

gaagcacatc tgttagtgac tcccagcagg ccatggacta gtcactaacc ccacactcaa 3780
aggggcatgg gtgggtggaga agcagaaggc gcaatcaagc ttatctggat atttctttct 3840
ttatttattt tacatggaaa taatatgatt tcacttttc tttagtttct ttgctctacg 3900
tgggcacctg gcactaaggg agtaccttat ttcctcacat cgcggaaatttc aacagctaca 3960
ttatattcc ttctgacact tggaaggat tggaaatttctt agaaaatgtat ctttctcaca 4020
aagttagagac caagagaaaa actcattgat tgggtttcta ctctttcaa ggactcagga 4080
aatttcactt tgaactgagg ccaagtgagc tgtaaagata accccacactt aaactaaagg 4140
ctaagaatat aggcttgatg gggaaattgaa ggttaggctga gtattggaa tccaaattga 4200
attttgattc tccttggcag tgaactactt tgaagaagtg gtcaatgggt ttttgcgtcc 4260
atgagcatgt acaacctctg gagctagaag ctcctcagga aagccagttc tccaaaggatct 4320
taacctgtgg cactgaaaagg aatgttgatg taccttctca ttttttagac agccaaaccct 4380
atccattaaa gtacttggta gaacacaaaa aaaaaaaaaa agggcgg 4427

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)